



TRABAJO DE FIN DE MÁSTER

MÉTODOS ESTADÍSTICOS PARA EVALUAR LA CAUSALIDAD EN ESTUDIOS OBSERVACIONALES

JUNIO 2017

Máster en Bioestadística

Facultad de Estudios Estadísticos

Universidad Complutense de Madrid

CLAUDIA COSCIA REQUENA

TUTORES: ALFONSO MURIEL GARCÍA

y TERESA PÉREZ PÉREZ

Agradecimientos:

A toda la Unidad de Bioestadística Clínica del Hospital Universitario Ramón y Cajal, por darme la oportunidad de formarme con ellos durante mi estancia, enseñándome y transmitiéndome su pasión por el trabajo y por haber confiado en mí.

Gracias también a los médicos de la UCI del Hospital Universitario de Getafe, en especial a Oscar Peñuelas, por haberme puesto a disposición la base de datos con la que he trabajado y haberme ayudado a establecer las variables clínicas relevantes del estudio.

A Alfonso y Teresa, por su ayuda constante y su interés en que el trabajo saliera lo mejor posible.

A mi familia.

Índice de Contenidos

Índice de Tablas.....	- 5 -
Índice de Figuras	- 5 -
Resumen.....	- 6 -
1. Introducción	- 7 -
2. Métodos	- 9 -
2.1. Índice de Propensión.....	- 12 -
2.2. Métodos del IP	- 14 -
2.2.1. Emparejamiento (PS matching).....	- 14 -
2.2.2. IPTW	- 19 -
2.2.3. Estratificación	- 20 -
2.2.4. Ajustes analíticos.....	- 20 -
2.3. Selección de las variables que incluir en el IP	- 20 -
Resumen, ventajas e inconvenientes de los cuatro métodos.....	- 21 -
2.4. IP Versus Regresión logística	- 23 -
3. Descripción Base de datos	- 25 -
4. Aplicación sobre un escenario clínico I	- 27 -
4.1. Búsqueda de variables confusoras.....	- 28 -
4.2. Estimación y métodos del Índice de Propensión	- 31 -
4.2.1. Emparejamiento.....	- 31 -
4.2.2. IPTW	- 37 -
4.2.3. Estratificación.....	- 38 -
4.2.4. Ajustar el IP como covariable en el modelo.....	- 39 -
4.3. Estimación del efecto sin el índice de propensión.....	- 40 -
4.4. Conclusiones sobre el escenario clínico I	- 43 -
5. Aplicación sobre un escenario clínico II	- 45 -
5.1. Búsqueda de variables confusoras.....	- 45 -
5.2. Estimación y métodos del Índice de Propensión (PS).....	- 48 -
5.2.1. Emparejamiento.....	- 48 -
5.2.2. IPTW	- 53 -
5.2.3. Estratificación.....	- 54 -
5.2.4. Incluyendo el PS como covariable del modelo.....	- 55 -
5.3. Estimación del efecto sin el índice de propensión.....	- 55 -
5.4. Conclusiones sobre el escenario clínico II	- 57 -
6. Conclusiones.....	- 58 -
7. Bibliografía	- 60 -

Índice de Tablas

Tabla 1. Diferencias entre EO y EC.....	- 8 -
Tabla 2 Ventajas e inconvenientes de los cuatro métodos propuestos del IP.	- 23 -
Tabla 3 Estudio de las variables confusoras para el evento en el escenario clínico I.....	- 29 -
Tabla 4 Estudio de las variables confusoras para la exposición en el escenario clínico I.	- 30 -
Tabla 5 Comparación de las variables basales entre tratados y no tratados antes y después del emparejamiento en el escenario clínico I	- 36 -
Tabla 6 Resumen de los distintos métodos utilizado para el escenario clínico I s	- 43 -
Tabla 7 Estudio de las variables confusoras para el evento en el escenario clínico II	- 47 -
Tabla 8 Estudio de las variables confusoras para el tratamiento en el escenario clínico II....	- 48 -
Tabla 9. Comparación de las variables basales entre tratados y no tratados antes y después del emparejamiento en el escenario clínico II	- 52 -
Tabla 10 Resumen de los distintos métodos utilizados en el escenario clínico II	- 57 -

Índice de Figuras

Figura 1 Gráfico causal entre tratamiento, evento y variables confusoras	- 8 -
Figura 2 Relación de pacientes expuestos y no expuestos	- 9 -
Figura 3 Esquema del método de emparejamiento	- 15 -
Figura 4 Zona de soporte común según gráfico de cajas	- 18 -
Figura 5 Zona de soporte común considerando la función de densidad	- 18 -
Figura 6 Relación causal entre NMB, Delirio y variables confusoras	- 27 -
Figura 7 Zona de soporte común en el escenario clínico I.....	- 33 -
Figura 8 Relación causal entre la interrupción de la sedación, mortalidad y variables confusoras	- 45 -
Figura 9 Zona de soporte común en el Escenario Clínico II	- 50 -

Resumen

Antecedentes: En investigación clínica es frecuente que uno de los objetivos sea establecer la asociación causal que existe entre una exposición y un desenlace. En los estudios observacionales, si la exposición está relacionada con las características del individuo, los sujetos expuestos y no expuestos difieren, además de por la exposición recibida, por otras características. Los efectos estimados por diferentes métodos estadísticos difieren según el escenario teórico establecido.

Objetivo principal:

Comparar los efectos estimados por diferentes métodos estadísticos para evaluar la causalidad en estudios observacionales en dos escenarios que presentan frecuencias diferentes en los eventos, en las exposiciones y en el número de variables a considerar en los ajustes.

Métodos

Análisis estadístico: Se plantean modelos de regresión logística univariable y multivariable ajustando por variables confusoras. Se utiliza el Índice de propensión (IP) y se estima el efecto mediante distintos métodos del IP.

Resultados:

Si los sujetos expuestos y no expuestos no se diferencian, las estimaciones por los distintos métodos coinciden. Cuando el número de eventos y de exposiciones es reducido, el efecto del NMB en el delirio varía en términos de OR.

Conclusiones

Las características del escenario teórico subyacente deben ser tenidas en cuenta para seleccionar el método más adecuado de ajuste en los análisis de causalidad. El índice de propensión permite seleccionar adecuadamente variables por las que corregir.

Si existen diferencias en los perfiles de exposición, la adecuación de los distintos métodos depende de la prevalencia de las exposiciones y de la incidencia de los desenlaces.

Palabras clave: estudio observacional, confusión, índice de propensión, regresión logística

1. Introducción

En investigación clínica, uno de los objetivos principales es evaluar la asociación causal que existe entre un tratamiento (exposición) y el desenlace de salud de un paciente.

El ensayo clínico [1] es un experimento prospectivo con el objetivo de evaluar intervenciones médicas, sanitarias, o quirúrgicas con una pregunta claramente definida en el que la asignación de los tratamientos se realiza dentro de un proceso aleatorio.

Los ensayos clínicos aleatorizados (ECA) son el tipo de investigación de referencia (gold-estándar) cuando se quiere estimar el efecto que produce el tratamiento sobre el evento de interés, ya que, debido a su condición de *ensayo*, [2] se obtienen estimaciones insesgadas y se reduce el error de estimación. En los ECA los distintos tratamientos disponibles se asignan aleatoriamente a los individuos que constituyen la muestra. Esta distribución permite que, covariables tan importantes como el estadio de la enfermedad, estén equilibradas entre los distintos tratamientos y se controlen los sesgos de los resultados. De esta forma, se podrá estimar el efecto de modo directo.

Aunque se consideren los estudios con máxima evidencia para la estimación de la causalidad, los ECA también presentan algunas limitaciones: los criterios de inclusión reducen el marco de la población a analizar, ya que tienden a excluir pacientes con edades avanzadas y con más de una enfermedad, mujeres embarazadas, niños, etc. Además, los ECA presentan aspectos éticos a tener en cuenta y el tiempo de seguimiento es reducido. Para poderse llevar a cabo, necesitan un grupo control sobre el cual medir la diferencia entre tratamientos. Este grupo, no siempre coincide con el establecido en la práctica clínica.

Una posible solución a los ECA, son los estudios observacionales (EO). En estas investigaciones no hay ninguna intervención por parte del investigador ni se lleva a cabo una asignación aleatoria del tratamiento [1], estos vienen determinados en varias ocasiones por las características clínicas del propio paciente y la práctica habitual del médico.

Los EO presentan varias ventajas en comparación con los ECA: el marco de la población es más amplio, así como el tiempo de seguimiento y pueden presentar un tamaño

muestral mayor. Las principales diferencias entre ambos estudios se pueden observar en la *Tabla 1*:

Estudios Observacionales	Ensayos clínicos
<ul style="list-style-type: none"> • Ventajas <ul style="list-style-type: none"> • Fáciles de realizar • Mayor reclutamiento de pacientes • Coste más bajo • Mayor representación de la población (validación externa) • Menor riesgo de los pacientes • Inconvenientes <ul style="list-style-type: none"> • Menor evidencia de la relación causa efecto 	<ul style="list-style-type: none"> • Ventajas <ul style="list-style-type: none"> • Minimización de la confusión • Permite pruebas de intervención • Debido a la evidencia clínica, presenta una inferencia causal mejor • Mayor impacto de publicación • Inconvenientes <ul style="list-style-type: none"> • Más difícil de llevar a cabo • Menos representación de la población (validación externa) • Riesgo inherente del estudio

Tabla 1. Diferencias entre EO y EC

Por otra parte, el hecho de que en un EO la asignación del tratamiento no sea aleatoria, conlleva que el concepto de causalidad no sea apropiado, debido en parte, a que pacientes tratados y no tratados difieran no solo en el tratamiento que reciben sino también en sus características basales. Si estas variables tuvieran asociación con el evento de interés, podrían actuar como variables confusoras entre la exposición y el resultado. El esquema de casualidad se puede observar en la *Figura 1*:



Figura 1 Gráfico causal entre tratamiento, evento y variables confusoras

Por lo tanto, el objetivo principal de este Trabajo de Fin de Máster es identificar y comparar dos métodos estadísticos utilizados para evaluar la causalidad en estudios observacionales, en concreto, el índice de propensión y la regresión logística aplicados en diferentes escenarios clínicos.

Otros aspectos que se van a tratar a lo largo del TFM es una comparativa de las distintas modalidades que presenta el propio índice de propensión.

En resumen, este trabajo constará principalmente de dos partes: en el apartado 2 de *Métodos* se explican y se detallan las características del método estadístico “propensity score” y sus técnicas de aplicación. Además, se indican los varios pasos que se tienen que realizar para estimar el efecto del tratamiento sobre el evento de interés. Por último, se explican de modo teórico la diferencia entre el índice de propensión y la regresión logística.

Por otro lado, en los apartados 4 y 5 se ponen en práctica todos los aspectos teóricos descritos anteriormente mediante el uso de varios escenarios clínicos distintos donde la prevalencia de la exposición y del tratamiento influyan en la estimación obtenida entre ambos métodos.

Por último, en el apartado 7 se comparan los resultados obtenidos con los distintos métodos, tanto a nivel metodológico como a nivel clínico y se presentan las conclusiones pertinentes.

2. Métodos

En un estudio observacional para evaluar el efecto de un tratamiento es necesario tener dos grupos de pacientes: aquellos que recibieron el tratamiento (pacientes expuestos) y pacientes que no recibieron el tratamiento (no expuestos), como se observa en la Figura 2:

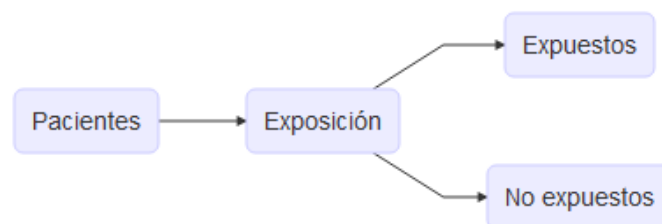


Figura 2 Relación de pacientes expuestos y no expuestos

En los EO, no hay una asignación aleatoria del tratamiento al paciente, sino que es el investigador/clínico el que habrá elegido previamente qué tratamiento asignar, en la mayoría de los casos, basándose en las características basales del paciente.

Por eso mismo, los dos grupos formados no sólo difieren en el tratamiento recibido sino también en las características clínicas y demográficas que presentan.

Como detallan Rosenbaum y Rubin en 1983 [3], en un ECA se puede estimar el efecto de una exposición de modo directo puesto que pacientes tratados y no tratados son similares.

Por otra parte, en los estudios no aleatorizados, los pacientes tratados suelen ser diferentes que los no tratados. Esta diferencia entre grupos conlleva que la estimación de la causalidad de la exposición sobre el desenlace de salud sea sesgada (dependa de las condiciones basales) y no se pueda obtener directamente. Las características basales de los pacientes son las responsables de que a un paciente se le asigne un tratamiento u otro, por lo que estas variables se podrían considerar variables confusoras: las condiciones basales de los pacientes conllevan que el clínico opte un tratamiento u otro por lo que el efecto que se obtiene con el tratamiento también dependerá de las condiciones del paciente y, en consecuencia, se vería confundido.

Haciendo referencia a la *Figura 1*, la confusión se detecta cuando la asociación entre dos variables (exposición-evento) difiere considerablemente cuando se considere, o no, otra variable (confusora). En este caso en concreto, las variables basales de los pacientes serán a menudo las variables confusoras.

Por lo tanto, para poder estimar los efectos de una exposición sobre un desenlace de salud en los EO, el primer paso debe ser corregir los problemas de confusión. Para ello, lo primero que se debe hacer es identificar la(s) variable(s) confusora(s) del estudio.

Las variables confusoras son aquellas que producen un cambio en el efecto de la exposición sobre el evento, matemáticamente es habitual establecer confusión si el cambio en el efecto de interés es al menos un 10%. En la práctica, muchas veces se identifican estas variables a partir de criterios clínicos bien establecidos.

Una vez que se ha encontrado las posibles variables confusoras hay distintos métodos estadísticos para resolver el problema [4].

- **Emparejar:** Es un método que consiste en realizar parejas de individuos con características similares. Fijando un individuo con tratamiento, se busca otro individuo sin tratamiento pero que presente características similares y se comparan las proporciones de evento en cada grupo. No es muy sencillo de aplicar en la práctica puesto que se debe buscar un tratado y un no tratado lo

más similar al tratado. Otra variante de esta opción sería emparejar por lo eventos (diseño de casos y controles).

- **Ajustar por confusores (mediante regresión):** Se estima un modelo de regresión logística para evaluar el efecto del tratamiento en el desenlace, en el cual las variables independientes son las confusoras junto con el tratamiento, y la variable respuesta es el evento que se quiere estudiar. Este método permite ajustar por distintas variables confusoras.
- **Ajustar por confusores (estratificación):** Este método consiste en fijar los niveles (estratos) de la variable confusora y recrear grupos dentro de dichos estratos para que no haya confusión: dentro de cada estrato obtenido se considera que los pacientes presentan características parecidas. Por lo tanto, se estima el efecto del tratamiento dentro de cada factor. Cuando hay más de dos variables confusoras es prácticamente imposible de llevar a cabo puesto que se debería estratificar por cada estrato de cada variable confusora.
- **Mediante el Índice de Propensión (IP):** Consta de dos partes. En la primera parte se estima el índice de propensión, definido como la probabilidad de que un individuo reciba un tratamiento u otro, en base a sus características basales. Se calcula mediante un modelo de regresión logística, donde la variable dependiente es el tratamiento y las independientes son las confusoras. Una vez estimado el IP, éste se utiliza para estimar el efecto del tratamiento sobre el evento.

Cuando el problema de la confusión queda controlado, se debe estimar el efecto que produce el tratamiento sobre el evento, mediante las técnicas estadísticas de estimación apropiadas.

¿Cómo se mide el efecto del tratamiento?

Según Austin [5], teniendo en cuenta un conjunto de pacientes ($i=1\dots n$) y dos tratamientos posibles ($T=1$, $T=0$), cada individuo tendrá dos resultados potenciales: Y_i (1) es el resultado potencial que obtendría el individuo i si recibiera el tratamiento ($T=1$), e Y_i (0) es el resultado potencial que obtendría el individuo i si no recibiera el tratamiento ($T=0$).

Además, se dispone de la variable observada D :

$$\left\{ \begin{array}{l} D_i = 1, \text{ si el paciente } i \text{ ha recibido el tratamiento.} \\ D_i = 0, \text{ si el paciente } i \text{ no ha recibido el tratamiento.} \end{array} \right.$$

Por consiguiente, el conjunto de pacientes que reciben el tratamiento se les considerarán pacientes tratados (T) y el conjunto de individuos que no hayan recibido el tratamiento se considerarán pacientes no tratados (\bar{T}). En los estudios de causalidad, el individuo podrá recibir sólo un tratamiento y, por lo tanto, en la práctica se obtendrá un solo resultado potencial. La fórmula del resultado potencial es la siguiente:

$$Y_i = D_i Y_i(1) + (1 - D_i) Y_i(0)$$

Dependiendo del objetivo y de la naturaleza del problema, se podrá estimar el efecto de dos modos distintos:

- **Efecto medio del tratamiento (ATE):** $ATE = E[Y_i(1) - Y_i(0)]$

Es la diferencia de los resultados potenciales de un paciente.

- **Efecto medio del tratamiento en los tratados (ATT):**

$$ATT = E[Y_i(1) - Y_i(0) | D = 1]$$

Es la diferencia de los resultados potenciales de un paciente, considerando que efectivamente dicho paciente ha recibido el fármaco.

En los ensayos clínicos aleatorizados, los valores obtenidos entre el ATE y el ATT son los mismos ya que la aleatorización del tratamiento garantiza que pacientes tratados y no tratados sean iguales.

Por otro lado, en los estudios observacionales estas dos estimaciones difieren debido a que los pacientes de ambos grupos pueden presentar características diferentes. En estos casos, será el investigador el que, teniendo en cuenta sus objetivos, decidirá qué efectos son necesarios estimar.

2.1. Índice de Propensión

El Índice de Propensión (IP), o también Propensity Score (PS), según Rosenbaum y Rubin [3], se define como la probabilidad que presenta un paciente de ser asignado a un tratamiento, en función de sus características basales.

El IP es una probabilidad condicionada cuya fórmula es:

$$IP_i = P(D_i = 1 | X_i)$$

Donde X_i es el valor de las covariables para el individuo i -ésimo y D_i es una variable dicotómica que indica la exposición al tratamiento: tomará el valor 1 ($D_i=1$) si se ha asignado el tratamiento, y el valor 0 ($D_i=0$) si se ha asignado el control.

El índice de propensión cumple los criterios de un índice balanceado: la distribución de las variables basales medidas, condicionada al índice de propensión es similar entre tratados y no tratados. Es decir, en un conjunto de sujetos que tienen el mismo IP, la distribución de las variables observadas basales es la misma entre pacientes tratados y no tratados.

En estudios experimentales, como en un ensayo clínico aleatorizado, el índice de propensión es conocido, [3]: si el tratamiento presenta dos ramas (medicamento o placebo) se sabe que la probabilidad de que se asigne uno u otro es equiprobable y conocida, es decir, cada paciente tiene una probabilidad de 0,5 de que se le asigne cada rama (o bien el medicamento o bien el placebo). Sin embargo, en el caso de los estudios observacionales este índice es desconocido, puesto que al depender de las características basales es difícil, o casi imposible, predecir esa probabilidad. Mediante regresión logística se puede llegar a estimar el Índice de propensión.

Los métodos basados en el índice de propensión controlan el efecto de la confusión mediante la comparación de sujetos tratados y no tratados con distribución similar de las covariables iniciales.

Para estimar el índice de propensión de modo correcto, se tienen que cumplir dos condiciones:

- **No hay confusores no medidos.** Se conocen de antemano las variables confusoras y se tiene medido el efecto que producen. Dado que esta asunción no se puede comprobar, Rosenbaum y Rubin [3] propusieron realizar un análisis de sensibilidad para estimar el efecto de un confusor no medido para ver si los resultados se ven alterados.
- Cada sujeto tiene una probabilidad distinta de 0 de recibir el tratamiento.

Si se cumplen dichas condiciones, se podrá concluir que la asignación del tratamiento es independiente del resultado.

Para estimar el índice de propensión se utilizan modelos de regresión logística, donde la variable respuesta será la variable que en el estudio se considere como el tratamiento o

exposición y las variables explicativas serán las características basales de los individuos, entre las que se encontrarían las posibles variables confusoras.

Partiendo de un modelo de regresión logística básica, la fórmula del modelo es:

$$P(Y = 1) = \frac{1}{1 + e^{-(\alpha_0 + \alpha_1 x_1 + \dots + \alpha_n x_n)}}$$

Donde Y será la variable dependiente (tratamiento), y $x_1 \dots x_n$ las variables basales de los pacientes.

Para obtener el IP estimado, se deberá modelizar la regresión logística y obtener los valores predichos del modelo. Una vez que se tiene los valores del PS estimado, se aplicarán para evaluar el efecto de la exposición sobre el evento.

2.2. Métodos del IP

Existen distintos métodos de aplicación del índice de propensión. Según Austin [6], los más utilizados son 4 métodos: emparejamiento, estratificación, ponderar por la inversa del índice de propensión, y ajustar un modelo considerando el IP como covariable. A continuación, se describirán y se detallarán los pasos que se deben seguir en cada método para la obtención de la estimación.

2.2.1. Emparejamiento (PS matching)

El método del emparejamiento, o también conocido como Propensity Score matching, se basa en crear una nueva muestra de individuos creando parejas de pacientes que presenten valores estimados del IP similares. El objetivo de este método es crear una muestra que sea balanceada para reducir, en consecuencia, las diferencias que pueda haber entre pacientes tratados y no tratados.

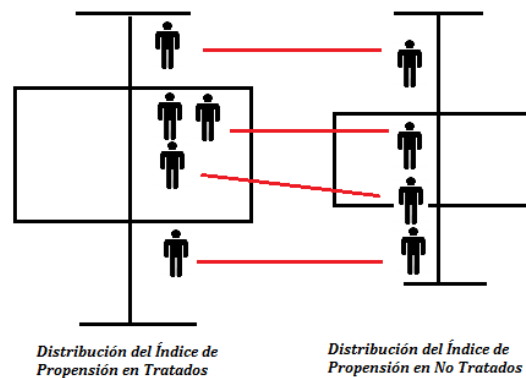


Figura 3. Esquema del método de emparejamiento

El método del emparejamiento presenta diferentes alternativas de búsqueda y selección de la pareja óptima.

- **Emparejamiento de forma codiciosa o “greedy”.** Consiste en seleccionar un paciente tratado de modo aleatorio y buscar el no tratado que presente el valor del IP más cercano. Este proceso se repite hasta que todos los pacientes tratados encuentren pareja. Durante todo el procedimiento *greedy*, no se está considerando que ese mismo paciente pudiera ser una mejor pareja para otro paciente.
- **Emparejamiento óptimo,** Consiste en seleccionar un paciente tratado de modo aleatorio y buscar el no tratado que presente el valor del IP en el cual las parejas están formadas con el fin de minimizar la diferencia total entre parejas. Al comparar con *greedy*, Austin [5] ha comprobado que el emparejamiento óptimo no produce mejores muestras balanceadas que el emparejamiento de modo codicioso.
- **Emparejamiento con el vecino más cercano,** de este modo se selecciona el paciente no tratado que tenga el valor más cercano del PS del paciente tratado. En el caso en el que hubiera más de un paciente no tratado con un PS similar, se deberá elegir de modo aleatorio.
- El método del emparejamiento permite fijar un calibre (*caliper*) que indique la distancia máxima con la que se puede emparejar a dos pacientes. Se suele utilizar como calibre el 25% de la desviación típica del PS estimado: $0.25 \cdot SD(PS)$ [7]
- Se puede considerar reemplazamiento, por lo tanto, se podrá decidir si un individuo puede encontrar pareja dos veces. En este caso se debe tener en

cuenta para la interpretación de la estimación de la varianza que un paciente no tratado podría estar en más de una pareja.

Independientemente del método de emparejamiento utilizado, la nueva muestra obtenida tendría una estructura muy similar a la de un ensayo clínico aleatorizado, Austin [8] comenta que las distribuciones de las covariables serían similares en ambos grupos formados. De este modo, se estaría solucionando el problema de la confusión y se estarían reduciendo las diferencias de las características de pacientes tratados y no tratados.

Se podría además estimar directamente el efecto del tratamiento comparando los resultados entre sujetos tratados y no tratados de la nueva muestra. Si la variable respuesta es continua se podrá estimar la diferencia entre los dos grupos mediante el test de la t de muestras relacionadas. Si la variable respuesta es dicotómica, se estimará el efecto mediante McNemar y en términos de OR, utilizando modelos GEE o Modelos Mixtos.

La utilización de métodos estadísticos que tengan en consideración la no independencia entre tratados y no tratados es necesaria dado que no se puede asumir independencia.

Evaluación de la comparabilidad de los grupos

Un aspecto importante al estudiar la causalidad en estudios observacionales es que la diferencia que presenten los individuos tratados y no tratados sea mínima, es decir, que la distribución de las variables en ambos grupos sea similar. En los ECA esta condición se cumple, pero en el caso de los estudios observacionales es muy poco probable que ambos grupos sean similares debido a los factores de confusión que pueda haber.

Por eso mismo, una vez que se ha realizado el emparejamiento, es necesario comparar los valores de las distribuciones de las características basales entre tratados y no tratados antes y después del emparejamiento.

El objetivo del índice de propensión y en consecuencia del emparejamiento, es reducir las diferencias que pueda haber entre tratados y no tratados producidas por los factores de confusión. Por lo tanto, si hubiera diferencias en la distribución de una variable después del emparejamiento podría concluirse que el IP está mal especificado.

Debido a que no se pueden utilizar contrastes de hipótesis porque son muy sensibles al tamaño muestral, se utilizan las diferencias estandarizadas. Si estas diferencias son mayores de un 10% significará que hay discrepancias en las distribuciones de las variables basales entre ambos grupos y que el propensity score no está especificado correctamente.

Las diferencias estandarizadas se expresan de la siguiente forma para las variables continuas:

$$d = \frac{\bar{x}_T - \bar{x}_{\bar{T}}}{\sqrt{\frac{s_T^2 + s_{\bar{T}}^2}{2}}}$$

Siendo $\bar{x}_T, \bar{x}_{\bar{T}}$ las medias de las variables en los tratados y en los no tratados, y $s_T^2, s_{\bar{T}}^2$ las varianzas de dichas variables.

Del mismo modo se definen las diferencias estandarizadas para variables dicotómicas:

$$d = \frac{(\hat{p}_T - \hat{p}_{\bar{T}})}{\sqrt{\frac{\hat{p}_T(1 - \hat{p}_T) + \hat{p}_{\bar{T}}(1 - \hat{p}_{\bar{T}})}{2}}}$$

Donde \hat{p}_T es la proporción de cada categoría de la variable confusora para los individuos tratados y $\hat{p}_{\bar{T}}$ es la proporción para individuos no tratados.

Zona de soporte común

La zona de soporte común es la zona delimitada entre el valor máximo del índice de propensión de los pacientes no tratados y el valor mínimo en los no tratados. Es el rango de valores del índice de propensión en el cual pacientes tratados y no tratados coinciden. Como se observa en la *Figura 3*, la zona de soporte común es la delimitada por las líneas rojas:

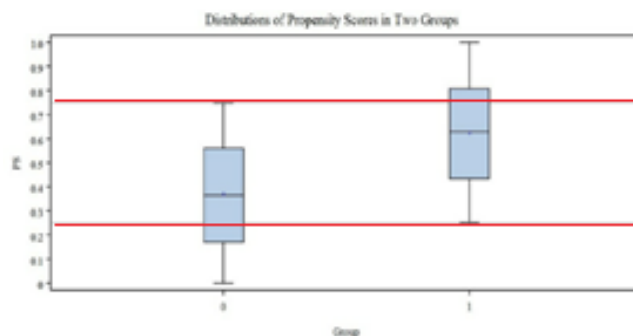


Figura 4 Zona de soporte común según gráfico de cajas ¹

A la hora de realizar el emparejamiento, se puede especificar considerar sólo la zona de soporte común para buscar las parejas: de este modo, las parejas que se creen serán muy similares. Considerar sólo la zona de soporte común conlleva también perder un número de pacientes que no se encuentren en ese rango.

Cuanto más parecidos sean los grupos de tratados y no tratados, la zona de soporte común abarcará mayor cantidad de pacientes de ambos grupos porque la distribución del índice de propensión en ambos grupos será parecida: presentará valores máximos y mínimos muy parecidos y por lo tanto la zona de soporte común incluirá a casi todos los pacientes.

Otro modo de ver la zona de soporte común que no sea mediante el gráfico de cajas, es mediante las funciones de densidad obtenidas para pacientes tratados y no tratados.

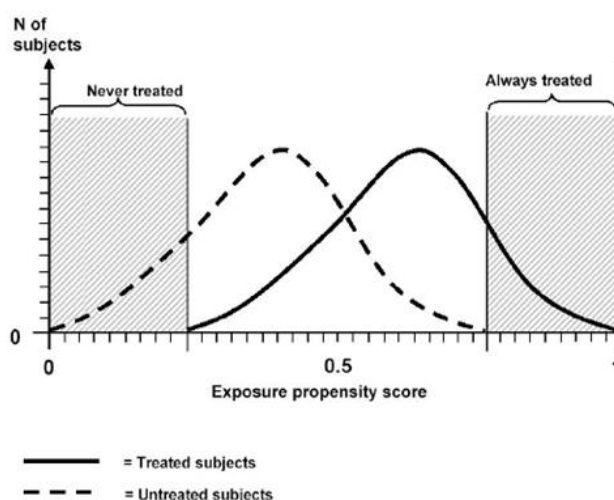


Figura 5 Zona de soporte común considerando la función de densidad ²

¹<http://cort.as/xNf2>

² <http://cort.as/xNf3>

De nuevo la zona de solapamiento de los pacientes tratados y no tratados es lo que se conoce como la zona de soporte común. Cuánto más parecidos sean pacientes tratados y no tratados, mayor será la zona de solapamiento. Si pacientes de ambos grupos fueran muy distintos no habría prácticamente zona de solapamiento.

2.2.2. IPTW

Otra aplicación del índice de propensión es Ponderando por la inversa de la probabilidad de tratamiento, según las siglas en inglés, IPTW: Inverse Probability Treatment Weighting.

Consiste en calcular pesos como la inversa del índice de propensión y posteriormente realizar la estimación del efecto considerando dichos pesos.

Considerando la variable D_i como aquella que indica si el paciente está o no tratado, se puede definir el peso del modo siguiente:

$$W_i = \frac{D_i}{IP_i} + \frac{1 - D_i}{1 - IP_i}$$

Por lo tanto, se obtiene que:

$$\left\{ \begin{array}{l} W_i = \frac{1}{IP_i} \text{ si Tratamiento} = 1 \\ W_i = \frac{1}{1 - IP_i} \text{ si Tratamiento} = 0 \end{array} \right.$$

Una vez que se han calculado los pesos, bastará con incluirlos en el modelo que se esté utilizando para la estimación del efecto.

Es decir, si se calculara el efecto mediante regresión logística, bastaría con especificar en el modelo que pondere por los pesos obtenidos.

Al ajustar por los pesos, la muestra del estudio sería representativa de una población específica.

Este método, junto con el emparejamiento, son los más utilizados dentro del índice de propensión.

2.2.3. Estratificación

El método de estratificación consiste en dividir la muestra en varias submuestras basadas en el índice de propensión de pacientes tratados y no tratados.

El número de grupos por los que se estratifica habitualmente es cinco, en función de los quintiles del índice de propensión estimado. Según Rosenbaum y Rubin [3], al estratificar la muestra en función de los quintiles, se reduce el sesgo en un 90%.

Cuando ya se tienen los estratos, se estimará el efecto del tratamiento mediante el OR crudo y mediante el OR de Mantel-Haenszel. Este último OR es el que corrige el efecto de las variables confusoras.

2.2.4. Ajustes analíticos

La última alternativa que se propone en este TFM es calcular el efecto del tratamiento mediante una regresión multivariable en la cual se incluya el índice de propensión como covariable confusora del modelo.

Este método tiene la premisa de que la relación entre el desenlace (evento) y la variable confusora deber estar correctamente especificada.

Se estimaría un modelo de regresión (logística o lineal, dependiendo de la naturaleza del evento) y se obtendrían las estimaciones pertinentes sobre el efecto. La interpretación del resultado estimado se realiza como en los modelos clásicos de regresión.

2.3. Selección de las variables a incluir en el IP

Las variables incluidas en el IP pueden crear sesgos, modificando la varianza del estimador o el error si no están correctamente elegidas. Para seleccionar dichas variables, no existe un consenso ya que depende del planteamiento clínico, por lo tanto, esta decisión dependerá tanto de criterios clínicos como estadísticos [9].

Dado un conjunto de variables basales, se pueden clasificar en: variables que se relacionan con el evento, variables que se relacionan con la exposición y variables que se asocian tanto con el evento como con la exposición.

En estudios de simulación, según Brookhart et al [10], se ha podido comprobar que las variables que se relacionan con el resultado (evento) se deben incluir, así como las variables que presenten asociación con el evento y la exposición. No se deberán incluir aquellas variables que presenten asociación sólo con el tratamiento.

Resumen, ventajas e inconvenientes de los cuatro métodos

Deb. et al (2016) realizan una comparación de los diferentes métodos del propensity score, siendo las principales diferencias las siguientes [11]:

De los cuatro métodos detallados, el emparejamiento es aquel que reduce más el sesgo, sobre todo comparado con el método de estratificación y del ajuste por covariable. Además, su presentación de los resultados es muy similar al de un ensayo clínico aleatorizado. Es relativamente sencillo de realizar e interpretar. Por otro lado, al no considerar en el estudio aquellos pacientes que no encuentran pareja se estaría disminuyendo la precisión del estudio y la posterior validez externa se ve modificada. Requiere por tanto de mayor número de pacientes no tratados que pacientes tratados.

El método de estratificación presenta la ventaja principal que utiliza todos los datos del estudio, además de permitir tanto la estimación del ATT como del ATE. Sin embargo, este método reduce menos el sesgo que otros métodos.

Ajustar el modelo con la covariable evita prescindir de pacientes del estudio y además permite incluir en una sola variable, un gran número de variables (el IP se crea mediante todas las variables basales). Por otro lado, se debe especificar correctamente la asociación entre el desenlace de salud y la variable confusora, el índice de propensión.

El modelo producido estimaría solo resultados en términos de OR y HR y no permite la estimación del efecto absoluto del tratamiento como puede ser la diferencia de riesgos y el NNT. Este método, además, se ha demostrado que produce estimaciones más sesgadas que otros métodos. Por último, con este método no es posible estimar el ATT ni el ATE.

Estimar el efecto del tratamiento utilizando la inversa de los pesos ponderados implica una mayor reducción del sesgo que los métodos de estratificación y el ajuste por covariable, además los resultados son muy similares al de un ECA y utiliza toda la base

de datos a disposición. Presenta la desventaja principal de que, al estar ponderando por unos pesos, aquellos pacientes que presenten un valor extremadamente bajo del índice de propensión obtendrán un peso muy alto por lo que un solo paciente podría ponderar mucho con relación al resto de pacientes: por ejemplo, un paciente con probabilidad de 0.000001 de ser asignado al tratamiento presenta un peso de 10^6 . Sin embargo, un paciente que presenta una probabilidad de asignación del tratamiento más alta, como, por ejemplo, del 30% (0.30), el peso por el que habría que ponderar sería de 3.33.

Un posible resumen de las ventajas e inconvenientes de los cuatro métodos del índice de propensión es el que aparece a continuación:

Métodos	Ventajas	Desventajas
<i>Emparejamiento</i>	Reducción mayor del sesgo en relación a estratificado y ajuste por covariable	Pacientes que no encuentran pareja son excluidos del estudio
	Presentación de resultados similar al ECA Fácil de realizar	Requiere mayor número de pacientes controles que tratados
<i>Estratificación</i>	Utiliza toda la base de datos	Reduce menos el sesgo que emparejamiento e IPTW No trabaja bien con análisis de supervivencia
<i>Ajuste por covariable</i>	Utiliza toda la base de datos	1 El modelo de regresión debe estar correctamente especificado. 2 Produce un sesgo mayor que otros métodos 3 La evaluación de la comparabilidad es más difícil que en otros métodos 4 No permite la estimación de la diferencia de riesgos
<i>IPTW</i>	Mayor reducción de sesgo que estratificación y ajuste por covariable Presentación de resultados similar al ECA Utiliza toda la base de datos	Más sensible a una posible mala especificación del IP y a valores extremos del IP

Tabla 2 Ventajas e inconvenientes de los cuatro métodos propuestos del IP.

2.4. IP Versus Regresión logística

La diferencia entre el índice de propensión y la regresión logística está muy condicionada al número de eventos y al número de covariables por las que ajustar [12].

Según Harrell (1996), el número óptimo de variables independientes que se debería incluir en un modelo de regresión logística es el número de eventos dividido por 10, lo que se conoce como la regla del 10 [13]. Es decir, si el número de eventos es 30 (fallecimientos, por ejemplo), el número óptimo de variables explicativas para estimar la probabilidad de fallecer debería ser 3. La regresión logística dependerá por lo tanto del número de eventos y en consecuencia del tamaño muestral.

Sin embargo, en la regresión logística del índice de propensión, la variable dependiente no es el evento, sino la exposición. El número de variables para estimar la exposición, no sólo dependerá de la asociación que presenten con el evento, como se decía anteriormente, sino también por el número de exposiciones que haya. Si, por ejemplo,

el número de exposiciones fuera 60 (60 fumadores), se debería ajustar el índice de propensión con seis variables.

Dependiendo de la prevalencia del evento y del número de expuestos al tratamiento, el índice de propensión permite ajustar un modelo con más variables que el modelo de regresión logística.

Bajo ciertas condiciones, las estimaciones obtenidas por los modelos de regresión multivariantes y el IP son coincidentes. La diferencia entre el índice de propensión y la regresión logística está muy condicionada al número de eventos y al número de covariables por las que ajustar [14].

3. Descripción Base de datos

La base de datos de la que se dispone para la realización de este Trabajo de Fin de Máster es una base de datos que reúne aquellos pacientes que ingresaron en Unidades de Cuidados Intensivos (UCI) de varios hospitales españoles, entre 1998 y 2016 en intervalos de seis años. Se recogieron datos durante los primeros 28 días de su estancia en la UCI. Por lo tanto, es un estudio prospectivo, multicéntrico y observacional.

Para la recopilación de las características de los pacientes, se midieron las variables de tipo *Cardinal 1*, es decir, aquellas variables que se midieron sólo una vez (variables basales), y el resto de variables que se fueron midiendo durante toda la estancia en la UCI. Se tendrán por lo tanto variables longitudinales, puesto que muchas de ellas se midieron en más de una ocasión.

Como se ha mencionado anteriormente, se tienen datos sobre distintos momentos de tiempo: años 1998, 2004, 2010 y 2016. No todas las variables aparecen en todos los años, por lo que para este TFM sólo se tendrán en cuenta los años 2010 y 2016, que son aquellos años que presentan las variables de interés, entre ellas, la variable que indica si el paciente ha recibido ventilación mecánica.

La base de datos con la que se trabajará presenta 2.335 observaciones y, considerando las variables longitudinales, se tienen en total alrededor de unas 30 variables.

Las variables principales que se utilizarán a lo largo de este TFM son:

- **NMB:** Variable dicotómica que indica si el paciente ha recibido bloqueantes neuromusculares.
- **Delirio:** Variable dicotómica que indica si el paciente ha presentado delirio a lo largo de la estancia en la UCI.
- **Status_day28:** Variable dicotómica que indica si el paciente ha fallecido o no durante a los 28 días de encontrarse en la UCI.
- **Daily_Awakening:** Variable dicotómica que indica si han interrumpido la sedación al paciente.
- **Edad:** Variable continua que indica la edad del paciente al ingreso en la UCI.
- **Sexo:** Variable dicotómica que hace referencia al sexo del paciente.

- **SAPS_II:** Variable continua que indica el grado de gravedad que presenta el paciente.
- **First ventilatory support:** Variable dicotómica que indica si el paciente ha recibido soporte de ventilación.
- **Cardiovascular failure:** Variable dicotómica que indica si el paciente ha presentado fracaso cardiovascular.
- **Renal failure:** Variable dicotómica que representa el fallo renal de un paciente.
- **Hematologic failure:** Variable dicotómica que indica si el paciente ha presentado fallo hematológico.
- **Sepsis_vm:** Variable dicotómica que indica si el paciente ha presentado sepsis durante la ventilación mecánica (vm).
- **Mvdays:** Variable continua que indica el número de días en el cual el paciente estuvo ingresado con ventilación mecánica.
- **Post-operative:** Variable dicotómica que indica si el paciente se encuentra hospitalizado en la UCI después de una operación quirúrgica.
- **Sepsis:** Variable dicotómica que indica si el paciente ha padecido sepsis.
- **Trauma:** Variable dicotómica que indica si el paciente ha presentado algún tipo de trauma.
- **Cardiac arrest:** Variable dicotómica que indica si el paciente ha presentado arresto cardíaco.
- **RR_total_day1:** Variable continua que indica la tasa de respiración que presenta el paciente (*Respiration Rate*) el primer día en la UCI.
- **Tidal_Volume_day1:** Variable continua que indica el valor del indicador *Tidal* en el primer día de ingreso en la UCI.

4. Aplicación sobre un escenario clínico I

A continuación, se muestra el uso y la aplicación del índice de propensión sobre la base de datos explicada en el apartado 3, para estudiar la interpretación tanto metodológica como clínica de los resultados obtenidos. Además, se aportarán los comandos básicos necesarios para la aplicación del IP del paquete estadístico STATA/IC 14.2.

En este primer escenario clínico, se quiere evaluar la causalidad que presentan los bloqueantes neuromusculares (NMB) sobre el delirio, en pacientes ingresados en UCI y que presentan enfermedades respiratorias, cuyo efecto puede verse confundido con ciertas características basales que presentan los pacientes.

En este caso no se tendrán a disposición el total de pacientes de la base de datos puesto que no todos ellos presentan enfermedades respiratorias. Para obtener sólo aquellos individuos que son de interés para este escenario clínico, se ha segmentado la muestra en relación a las variables que indican que el paciente presenta enfermedad respiratoria: tener asma, haber tenido una neumonía, alguna insuficiencia cardíaca congestiva (CHF) o tener síndrome de dificultad respiratoria aguda (ARDS).

El esquema de la causalidad sería el siguiente:

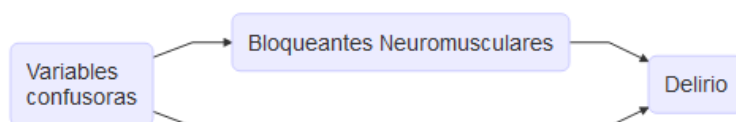


Figura 6 Relación causal entre NMB, Delirio y variables confusoras

Se observa que las variables confusoras presentan asociación tanto con los bloqueantes neuromusculares como con el delirio del paciente.

La particularidad de este escenario reside en la escasez de los eventos, así como en el número de tratamientos.

Mediante una tabla cruzada y el test de independencia de la chi-cuadrado se estudia la asociación entre ambas variables:

```
tab Delirium NMB, chi
```

Delirium	NMB		Total
	0	1	
0	377	135	512
1	13	13	26
Total	390	148	538

Pearson chi2(1) = 6.9299 Pr = 0.008

El p-valor asociado al contraste de la chi-cuadrado es de 0.008. Siendo este menor que el nivel de significación α prefijado (0.05) y la hipótesis nula es la independencia entre ambas variables, se concluye que las dos variables presentan asociación entre ellas.

Se observa que en este escenario clínico se tienen a disposición pocos eventos (sólo 26 pacientes presentan finalmente el delirio) y además se tienen pocos pacientes expuestos (sólo 148 han recibido bloqueantes neuromusculares).

En el ejemplo que se presenta a continuación, se considerarán confusoras aquellas variables basales que tengan asociación con el evento, con la exposición y que además el clínico/investigador haya considerado como clínicamente relevantes.

4.1. Búsqueda de variables confusoras

Lo primero que se debe hacer es determinar las variables confusoras que presenta el modelo, es decir, aquellas variables que muestren asociación con el evento y con la exposición.

Para ello, se deberá estudiar la asociación que cada una de las variables basales (posibles confusoras) presente con las categorías de las variables de interés: tanto el evento como el tratamiento.

Dicha asociación se estudiará con el contraste Chi-Cuadrado para las variables dicotómicas y mediante el test de Mann-Whitney para variables continuas. Ambos contrastes de hipótesis se mostrarán en la tabla con el p-valor asociado al estadístico del contraste. Además, se presentará, para las variables dicotómicas, la frecuencia relativa de los pacientes que presentan dicha característica acompañada del valor en % y las variables continuas se mostrarán con la media \pm la desviación típica. Se tomará como nivel de significación $\alpha = 0.05$.

Por eso mismo se crean la tabla 3 y la tabla 4: estas tablas recogen la información acerca de las variables basales y su asociación con el evento y el tratamiento.

Como se ha comentado anteriormente, se considerarán variables confusoras aquellas que presenten asociación con el evento y con el tratamiento.

- **Tabla de estudio de variables confusoras para el evento:**

	Si Delirium (n=27)	No Delirium (n=516)	P-valor
Edad	57,04 ± 14,50	64,10 ± 13,64	0,011
Sexo (Masculino)	6 (22,22)	268 (51,94)	0,003
IMC	28,20 ± 6,92	27,47 ± 6,50	0,630
Traqueotomía previa (Si)	1 (3,70)	11 (2,13)	0,588
SAPS_II	45,11 ± 16,31	45,95 ± 16,48	0,820
First Ventilatory Support (Non-invasive)	8 (29,63)	110 (21,32)	0,307
ICU_pneumonia (Si)	4 (22,22)	17 (5,03)	0,003
sepsis_vm (Si)	15 (55,56)	143 (27,71)	0,002
cardiovascular_failure (Si)	19 (70,37)	243 (47,09)	0,018
renal_failure (Si)	11 (40,74)	137 (26,55)	0,106
hepatic_failure (Si)	21 (77,78)	119 (23,06)	0,000
hematologic_failure (Si)	7 (25,93)	65 (12,60)	0,047
barotrauma (Si)	0 (0)	16 (3,10)	0,353
RR_total_day1	21 ± 5,75	19,89 ± 5,13	0,272
respiratoryRateVent1	20,89 ± 5,95	19,91 ± 5,20	0,602
Tidal_Volume_day1	468,56 ± 68,49	485,16 ± 90,72	0,296
Tidal_mlk_g_day1	6,12 ± 1,47	6,65 ± 1,81	0,143
Tidal_mlk_gPBW_day1	7,62 ± 1,13	8,10 ± 1,71	0,215
peakPressure_day1	30,04 ± 7,83	28,71 ± 7,98	0,520
plateauPressure_day1	20,79 ± 5,18	21,60 ± 6,24	0,532
appliedPEEP_day1	7,34 ± 3,33	7,80 ± 3,16	0,599
driving_pressure_day1	13,78 ± 3,79	13,86 ± 5,40	0,999
Mvdays	12,72 ± 8,67	10,14 ± 14,98	0,009

Tabla 3 Estudio de las variables confusoras para el evento en el escenario clínico I

- **Tabla de estudio de variables confusoras para el tratamiento:**

	Si NMB (n=148)	No NMB (n=390)	P-valor
Edad	59,53 ± 13,53	65,33 ± 13,59	0,000
Sexo (Masculino)	84 (56,76)	189 (48,46)	0,086
IMC	27,17 ± 7,15	27,60 ± 6,26	0,224
Traqueotomía previa (Si)	4 (2,70)	8 (2,05)	0,648
SAPS_II	43,74 ± 15,50	46,82 ± 16,76	0,082

	Si NMB (n=148)	No NMB (n=390)	P-valor
First Ventilatory Support (Non-invasive)	32 (21,62)	83 (21,28)	0,932
ICU_pneumonia (Si)	16 (17,20)	5 (1,92)	0,000
sepsis_vm (Si)	64 (43,24)	94 (24,10)	0,000
cardiovascular_failure (Si)	102 (68,92)	159 (40,77)	0,000
renal_failure (Si)	63 (42,57)	83 (21,28)	0,000
hepatic_failure (Si)	61 (41,22)	77 (19,74)	0,001
hematologic_failure (Si)	36 (24,32)	35 (8,979)	0,000
barotrauma (Si)	13 (8,78)	3 (0,77)	0,000
RR_total_day1	21,43 ± 5,55	19,35 ± 4,89	0,000
respiratoryRateVent1	22,60 ± 5,45	18,08 ± 4,23	0,000
Tidal_Volume_day1	472,90 ± 101,88	488,89 ± 84,02	0,009
Tidal_mkg_day1	6,54 ± 1,94	6,65 ± 1,73	0,295
Tidal_mkgPBW_day1	7,76 ± 1,75	8,20 ± 1,65	0,003
peakPressure_day1	31,17 ± 7,38	27,81 ± 8,01	0,000
plateauPressure_day1	24,46 ± 5,78	20,35 ± 5,94	0,000
appliedPEEP_day1	9,06 ± 3,65	7,26 ± 2,80	0,000
driving_pressure_day1	15,67 ± 5,23	13,11 ± 5,17	0,001
Mvdays	15,9 ± 13,14	8,21 ± 14,82	0,000

Tabla 4 Estudio de las variables confusoras para la exposición en el escenario clínico I.

Aquellas variables que presenten un p-valor menor que 0.05 en la *Tabla 3* y que además coinciden con las de los de la *Tabla 4* serán las variables confusoras que se deberán utilizar para este apartado.

En este caso, se observa que las variables confusoras son: Edad, Sexo, sepsis_vm, ICU_pneumonia, cardiovascular_failure, hepatic_failure, hematologic_failure y mvdays, cuyo significado ya está explicado en el apartado 3. Esto significa que estas variables confunden el efecto que producen los bloqueantes neuromusculares en el delirio sobre aquellos pacientes ingresados en la UCI con enfermedades respiratorias.

Como se señaló anteriormente, la selección de variables es crucial para estimar correctamente el efecto de un tratamiento. En el caso de los estudios clínicos, la colaboración por parte del clínico es fundamental puesto que él determina posibles variables confusoras que pueda presentar el modelo.

4.2. Estimación y métodos del Índice de Propensión

4.2.1. Emparejamiento

Como se ha explicado en la Introducción, una de las aplicaciones más utilizadas del Índice de Propensión es el emparejamiento.

En este TFM se han considerado opciones del IP como: el vecino más cercano, sin reemplazamiento (un individuo sólo encontrará pareja una vez) y como caliper se utilizará el 25% de la desviación típica obtenida del PS. Además, para modelizar y estimar el índice de propensión, se utilizará un modelo de regresión logística.

El programa estadístico STATA, mediante uno de sus comandos, **psmatch2**, permite realizar la estimación del PS y el emparejamiento de modo directo.

Las variables que se han incluido son aquellas que en las Tablas 3 y 4 presentaban asociación con el evento.

El comando:

```
psmatch2 NMB age Sex cardiovascular_failure SAPS_II hematologic_failure
renal_failure sepsis_vm First_ventilatory_support mvdays, outcome(Delirium)
caliper(.046) neighbor(1) noreplacement logit
```

En este comando hay que indicar en primera posición la variable que representa el tratamiento que se está estudiando (NMB) y a continuación hay que incluir todas las variables confusoras del modelo. Mediante la opción `outcome(Delirium)` se indica cuál es el desenlace de salud de interés, en este caso, el delirio del paciente. La opción `neighbor(1)` indica que se deberá coger el vecino más cercano y mediante la opción `noreplacement` se indica que el individuo podrá encontrar pareja sólo una vez. El comando `psmatch2` además permite decidir con qué modelo estimar el IP, es decir, o bien con el modelo logit o probit. Por último, se ha definido como caliper el valor 0.046, resultado de estudiar la variable que se autogenera por defecto, `_pscore`, que recoge todos los valores del índice de propensión estimado:

```
summ _pscore
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_pscore	522	.2701149	.1829924	.0365052	.9080531

Y se obtiene: $caliper = 0,25 \times 0,183 = 0,0457$

El comando `psmatch2`, como se ha explicado anteriormente, estima de modo automático el índice de propensión. Por eso mismo, la primera salida son los resultados obtenidos de la estimación del modelo de regresión logística del índice de propensión. Dichos resultados coinciden con los que se obtendrían aplicando una regresión logística básica.

Cabe mencionar que, en este caso, no importará el nivel de significación que se obtiene, puesto que, como se lleva asumiendo durante todo el TFM, las variables confusoras son aquellas que presentan asociación con el evento y puede ocurrir que no presenten asociación con el tratamiento. Además, podría ocurrir que las variables que el clínico considere relevantes para el estudio no obtengan la significación estadística necesaria.

El modelo de regresión logística para la estimación del IP es el siguiente:

```
Logistic regression               Number of obs   =       522
                                LR chi2(9)        =       88.28
                                Prob > chi2       =       0.0000
Log likelihood = -260.38294       Pseudo R2    =       0.1449
```

	NMB	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
age		-.0227871	.0081292	-2.80	0.005	-.03872 -.0068543
Sex		-.2581207	.2195805	-1.18	0.240	-.6884906 .1722491
cardiovascular_failure		.8697593	.2550505	3.41	0.001	.3698695 1.369649
SAPS_II		-.0096561	.007294	-1.32	0.186	-.0239521 .0046398
hematologic_failure		.7097984	.2950099	2.41	0.016	.1315896 1.288007
renal_failure		.6502991	.243403	2.67	0.008	.173238 1.12736
sepsis_vm		.1480821	.2507986	0.59	0.555	-.3434741 .6396383
First_ventilatory_support		-.3214629	.2738982	-1.17	0.241	-.8582935 .2153677
mvdays		.0230086	.0080246	2.87	0.004	.0072806 .0387366
_cons		.1756673	.6237019	0.28	0.778	-1.046766 1.398101

There are observations with identical propensity score values.
The sort order of the data could affect your results.
Make sure that the sort order is random before calling `psmatch2`.

Una vez que el comando `psmatch2` ha estimado el índice de propensión, el siguiente paso es realizar el emparejamiento y a estimar el efecto que produce los NMB sobre el delirio de los pacientes, es decir, el ATT.

En el próximo cuadro se muestran los valores estimados antes (*unmatched*) y después de emparejar (ATT).

Se observa que antes del emparejamiento, la diferencia del efecto del tratamiento obtenido en pacientes tratados y controles era mayor (0.054) que después de haber realizado las parejas (0.033).

Variable	Sample	Treated		Controls	Difference	S.E.	T-stat
Delirium	Unmatched	.085	.031	.054	.020	2.61	
	ATT	.083	.05	.033	.032	1.03	

Como se explicaba en el apartado de Métodos, para realizar el emparejamiento, se pueden buscar las parejas delimitadas por la zona de soporte común (*common support region*).

psmatch2:	psmatch2: Common		
Treatment	support		
assignment	Off suppo	On suppor	Total
Untreated	0	381	381
Treated	21	120	141
Total	21	501	522

Se observa que, de los 522 pacientes del estudio, 21 pacientes tratados se quedarían fuera de la zona de soporte común: los valores que presentan del IP son mayores que el valor máximo del índice de propensión de los no tratados.

Considerando que hay solo 120 tratados que entran dentro de la zona de soporte común, sólo podrá haber 120 parejas. Se tendrán que descartar, por lo tanto, el resto de pacientes no tratados.

Mediante el comando: `graph box _ps, by(_treated)`, se puede obtener el gráfico de la zona de soporte común:

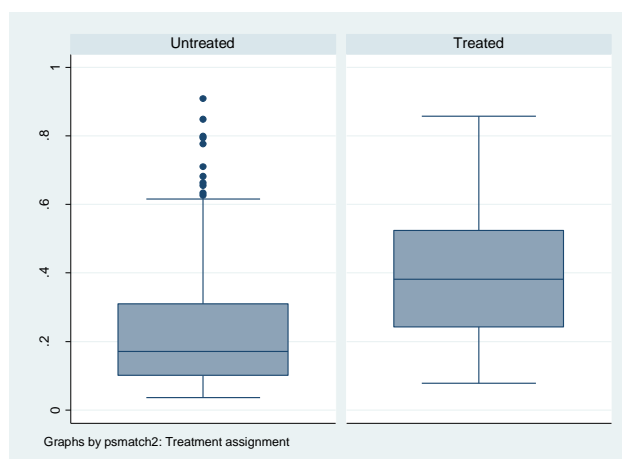


Figura 7 Zona de soporte común del escenario clínico I

Como se observa en la *Figura 7*, el comportamiento del índice de propensión en los dos grupos es diferente. Los pacientes que no han recibido bloqueantes neuromusculares (NMB) presentan, en media, valores del índice de propensión más bajos. Por otro lado, aquellos pacientes que sí han recibido NMB presentan valores medios más altos del PS.

Para estudiar de modo más específico los valores del índice de propensión, se puede realizar un resumen estadístico básico, diferenciando entre tratados y no tratados, como aparece a continuación:

```
summ _pscore if _treated == 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_pscore	141	.3921124	.1866678	.0784199	.8580513

```
summ _pscore if _treated == 0
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_pscore	381	.2249663	.1597107	.0365052	.9080531

La media del IP que se obtiene en ambos grupos es diferente, por lo que se puede concluir que los pacientes tratados y no tratados no presentan rasgos similares. Si las estimaciones que se obtuvieran del IP fueran similares, los dos gráficos de cajas presentarían un comportamiento similar y se podría concluir por lo tanto que los dos grupos de pacientes no presentan diferencias en cuanto a las características basales.

Si ahora se consideran sólo aquellos pacientes que han encontrado parejas, se observan los siguientes resultados:

```
summ _pscore if _treated == 1 & _weight == 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_pscore	120	.3443607	.1570842	.0784199	.8307891

```
summ _pscore if _treated == 0 & _weight == 1
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
_pscore	120	.3509538	.1656109	.0783619	.8489809

Se observa que si se consideran sólo los pacientes que han encontrado pareja los valores del índice de propensión en ambos grupos son muy similares.

Esto indica que, al emparejar mediante el índice de propensión, pacientes tratados y no tratados presentan características similares.

Una vez que se ha ejecutado el **psmatch2** y se han encontrado las mejores parejas según los criterios establecidos anteriormente, se debe calcular el efecto estimado de los NMB sobre el delirio. Para ello, al ser la variable de interés dicotómica, será suficiente con calcular el Odds Ratio (OR).

Con el comando:

cc Delirium NMB if _weight ==1, donde al seleccionar **_weight==1** se están seleccionando sólo aquellos pacientes que hayan encontrado pareja. Se obtiene la estimación del OR correspondiente:

	Exposed	Unexposed	Total	Proportion Exposed	
-----+-----					
Cases	10	6	16	0.6250	
Controls	110	114	224	0.4911	
-----+-----					
Total	120	120	240	0.5000	
	Point estimate		[95% Conf. Interval]		
-----+-----					
Odds ratio	1.727273		.5457754	5.974122	(exact)
Attr. frac. ex.	.4210526		-.8322554	.8326114	(exact)
Attr. frac. pop	.2631579				
-----+-----					
	chi2(1) =		1.07	Pr>chi2 = 0.3006	

El OR estimado es:

$$OR_{match} = 1.73, \text{ con } IC_{95}(OR_{match}) = [0.55, 5.97]$$

Este resultado significa que, si se utilizan NMB en pacientes con enfermedades respiratorias, la probabilidad de presentar delirio es un 73% mayor frente a aquellos pacientes a los que no se le han asignado los bloqueantes.

Una vez que se ha realizado la estimación del OR, el siguiente paso en el método del emparejamiento es realizar la evaluación de la comparabilidad de los grupos. Es decir, se debe comprobar que la distribución de las variables confusoras, antes y después del tratamiento sean similares y se hayan reducido las diferencias que podrían presentar. Este paso consiste en calcular las diferencias descritas en la *pág.13* y ver si éstas son menores que en alrededor un 10-15%.

Variables basales	Antes del emparejamiento			Después del emparejamiento		
	NMB = Si (n=148)	NMB = No (n=390)	Diferencias estandarizadas (%)	NMB = Si (n=120)	NMB = No (n=120)	Diferencias estandarizadas (%)
Edad (años)	59,53 ± 13,53	65,33 ± 13,58	42,79	60,53 ± 14,53	58,96 ± 13,64	11,17
SAPS_II	43,74 ± 15,49	46,83 ± 16,76	19,15	44,99 ± 15,67	43,41 ± 14,78	10,37
Sexo (V)	56,8	48,5	16,63	55	58,3	6,73
First_ventilatory_support (Si)	21,6	21,3	0,59	23,3	21,7	3,98
cardiovascular_failure (Si)	68,9	40,8	58,37	67,5	68,3	1,78
renal_failure (Si)	42,6	21,28	40,87	36,7	40	6,85
hematologic_failure (Si)	24,3	8,97	25,75	19,2	16,7	6,52
sepsis_vm	43,24	24,1	37	40,88	45,83	10,10
mvdays	15,90 ± 13,14	8,21 ± 14,82	54,91	14,35 ± 11,60	11,97 ± 19,04	15,09

Tabla 5 Comparación de las variables basales entre tratados y no tratados antes y después del emparejamiento en el escenario clínico I

Según la *Tabla 5*, se observa que las diferencias antes del emparejamiento eran más elevadas en comparación a después del emparejamiento. Por ejemplo, la variable edad presentaba una diferencia estandarizada de casi el 43% mientras que, al emparejar, los pacientes han reducido tratados y no tratados han reducido esta diferencia hasta un 11,17%.

Estos resultados significan que el índice de propensión, considerando las variables utilizadas, está bien definido. Si no fuera así, habría diferencias mayores. Se puede por lo tanto concluir que este modelo está especificado de modo correcto.

4.2.2. IPTW

A continuación, se procederá a aplicar el método IPTW para la estimación del efecto de los NMB sobre el delirio.

Cabe destacar que, si se hubiera estimado previamente el `_pscore` mediante el `psmatch`, como se ha hecho en este caso, dicha estimación valdría para realizar los siguientes cálculos y no haría falta estimarlo de nuevo.

Lo primero que se debe hacer es estimar los pesos asociados a cada individuo tratado y no tratado:

Una vez que se han calculado estos pesos, se debe estimar el modelo de regresión logística, incluyéndolos como ponderación. La variable dependiente será Delirium y la variable explicativa será NMB.

Para generar la variable `iptw` en Stata, se puede hacer mediante el siguiente comando:

```
gen iptw = 1 / _pscore if _treated==1
replace iptw= 1 / (1- _pscore) if _treated == 0
```

Se está indicando de este modo que, para los individuos tratados, el valor del `iptw` sea la inversa del IP y para los pacientes no tratados el valor del `iptw` sea la inversa de 1- el IP.

Para obtener la estimación del efecto, en este caso, simplemente es necesario añadir en el comando del `psmatch2` la condición de que realice las estimaciones ponderando por los pesos que se han calculado.

Al ser el delirio una variable dicotómica, para estimar el efecto, se debe estimar un modelo de regresión logística como aparece a continuación:

```
logistic Delirium NMB [pweight=iptw]
```

Mediante el comando **logistic** se está modelizando la probabilidad de presentar delirio si se han administrado bloqueantes neuromusculares, ponderado por los pesos obtenidos en el cálculo previo.

En este método, toda la información de las variables confusoras viene recogida mediante los pesos, por eso mismo no es necesario añadir ninguna otra variable que no sea el tratamiento.

El resultado es el siguiente:

```
Logistic regression                                Number of obs   =          522
                                                    Wald chi2(1)    =           2.43
                                                    Prob > chi2     =          0.1189
Log pseudolikelihood = -205.6152                Pseudo R2       =          0.0156
```

		Robust				
Delirium	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
NMB	2.083997	.9812445	1.56	0.119	.8281539	5.244246
_cons	.035942	.0110109	-10.86	0.000	.0197167	.0655193

Como se decía en el apartado 2.2.2. este método presenta la ventaja de que no se eliminan pacientes del estudio. En el método de emparejamiento el número de pacientes se reducía a 240 pacientes, aquí el modelo sigue presentando los 522 pacientes totales de la base de datos.

La interpretación de este resultado es igual a la de cualquier modelo de regresión logística.

El OR obtenido:

$$OR_{iptw} = 2.08, \text{ con } IC_{95}(OR_{iptw}) = [0.83, 5.24]$$

Este OR significa que es 2.08 veces más probable que un paciente al que han administrado bloqueantes neuromusculares presente un delirio durante su estancia en la UCI frente a aquellos pacientes que no han recibido NMB.

4.2.3. Estratificación

Para utilizar el método de estratificación, según la teoría explicada en el *apartado 2.2.3*, el primer paso que se debe hacer es crear la variable que especifique los quintiles del índice de propensión.

Una posible manera de calcular los quintiles y crear la nueva variable que se necesita es mediante el siguiente comando:

```
centile _pscore, centile( 20 40 60 80)

recode _pscore (min/.10199 =1 "Estrato 1") (.10199/.1684349=2 "Estrato
2") (.1684349/.2876358 =3 "Estrato 3") (.2876358/.4305894= 4 "Estrato
4") (.4305894/max = 5 "Estrato 5") , gen(ps_estr)
```

Una vez que se ha generado la nueva variable que representa los quintiles (*ps_estr*), se debe considerar para estimar el efecto. Mediante el comando *cc*, como se hacía en el apartado 4.2.1., se obtienen los resultados del OR, tanto Crudo como el de Mantel-Haenszel. Indicando la opción **by** (*ps_estr*), se está señalando que debe realizar la estratificación en relación a dicha variable.

cc NMB Delirium, by (ps_estr)

RECODE of _pscor	OR	[95% Conf. Interval]		M-H Weight	
Estrato 1	.	.	.	0	(exact)
Estrato 2	3.142857	.0497521	63.19744	.2666667	(exact)
Estrato 3	0	0	2.671533	1.105769	(exact)
Estrato 4	4.5	.6826716	48.84526	.6666667	(exact)
Estrato 5	1.72549	.3423984	11.22813	1.471154	(exact)
Crude	2.860465	1.141315	7.136999		(exact)
M-H combined	1.81655	.7457386	4.424948		

Test of homogeneity (Tarone) chi2(4) = 4.03 Pr>chi2 = 0.4023					
Test that combined OR = 1:					
Mantel-Haenszel chi2(1) = 1.71					
Pr>chi2 = 0.1908					

Los OR obtenidos son:

$$OR_{crudo} = 2.86, \text{ con } IC_{95}(OR_{crudo}) = [1.14, 7.14]$$

$$OR_{mantel} = 1.82, \text{ con } IC_{95}(OR_{mantel}) = [0.75, 4.42]$$

Las estimaciones obtenidas por ambos OR difieren entre ellas. Por lo tanto, mediante este método, se concluye que la probabilidad de presentar delirio es 1.82 veces mayor en pacientes que han recibido los bloqueantes neuromusculares frente a aquellos que no los han recibido.

4.2.4. Ajustar el IP como covariable en el modelo

El último método propuesto en el apartado 2.2.4 del IP es utilizar el índice de propensión como covariable en el modelo de regresión.

Este método es el más sencillo de realizar, puesto que consiste en introducirlo en el comando correspondiente de modo directo:

logistic Delirium NMB _pscore

Al añadir la variable `_pscore` en el comando, se están considerando todas las variables confusoras resumidas en una única variable.

La salida que se obtiene es la siguiente:

```

Logistic regression                                Number of obs   =          522
                                                    LR chi2(2)      =           8.83
                                                    Prob > chi2     =          0.0121
Log likelihood = -92.933498                        Pseudo R2       =          0.0454

```

Delirium	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]
NMB	2.068304	.9554359	1.57	0.116	.8363897 5.114697
_pscore	6.79998	7.554159	1.73	0.084	.7707263 59.99501
_cons	.020112	.0085961	-9.14	0.000	.0087025 .0464802

En este caso, el OR obtenido para la estimación del efecto es:

$$OR_{covariable} = 2.07, \text{ con } IC_{95}(OR_{covariable}) = [0.84, 5.11]$$

También en este caso, como ocurría en el resto de métodos del IP, se llega a la conclusión de que aquellos pacientes expuestos presentan una probabilidad mayor de presentar delirio frente a aquellos que no han recibido la exposición. En este caso, la probabilidad es dos veces mayor en pacientes tratados frente a los no tratados.

4.3. Estimación del efecto sin el índice de propensión

Uno de los objetivos de este TFM es comparar los resultados que se obtienen al aplicar el índice de propensión frente a la metodología clásica. En este caso, al no aplicar el IP, se debería estimar el efecto del NMB sobre el delirio mediante regresión logística, puesto que el evento es de naturaleza dicotómica (delirio o no delirio).

Recordando los valores que se tenían a disposición del evento y del tratamiento:

```

tab Delirium NMB, chi

```

	see notes		
Delirium	0	1	Total
0	377	135	512
1	13	13	26
Total	390	148	538

Se puede observar que el número de eventos de interés (Delirium=1) es 26 y el número de tratamientos (NMB=1) es 148.

Hay dos maneras de estimar el efecto sin el índice de propensión: considerando o no las variables confusoras del modelo.

Si no se estuvieran considerando las variables confusoras del modelo, se debería estimar un modelo de regresión logística univariable, en el cual la variable explicativa es la exposición o no de los NMB. Sin embargo, si se quisiera corregir por las variables confusoras, se debería estimar el efecto del NMB mediante un modelo de regresión logística multivariante, en el cual además de la exposición, se incluyen aquellas variables que se consideraron confusoras anteriormente.

Por lo tanto, si se aplicara la regla del 10 descrita en el apartado 2.4. para estimar el efecto que presentan los NMB frente al delirio en el caso de la regresión logística multivariable sólo se podría estimar mediante dos variables: $26/10 \sim 2$.

Por otro lado, en el caso del índice de propensión, debido a que la regresión logística se aplica para modelizar el tratamiento, se podría estimar con hasta 14 variables basales: $148/10 \sim 14$.

Para el modelo de regresión logística univariable solo se puede incluir el tratamiento como variable explicativa en la estimación del efecto. Se obtienen los siguientes resultados:

logistic Delirium NMB

Logistic regression	Number of obs	=	538
	LR chi2(1)	=	6.22
	Prob > chi2	=	0.0127
Log likelihood = -101.02743	Pseudo R2	=	0.0298

Delirium	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
NMB	2.792593	1.13059	2.54	0.011	1.262978	6.174748
_cons	.0344828	.0097273	-11.94	0.000	.0198374	.0599403

$$OR_{univ} = 2.79, \text{ con } IC_{95}(OR_{univ}) = [1.26, 6.17]$$

Este OR estimado indica que la probabilidad de presentar delirio en pacientes que han recibido el tratamiento es 2.79 veces mayor frente a aquellos pacientes no expuestos.

Por lo que respecta a la regresión logística multivariable, en este caso, al tener sólo 26 eventos, reduce el número de variables explicativas que incluir en el modelo a dos (incluyendo la exposición). Eso significa que se podrá ajustar sólo por una variable confusora.

Para elegir cuál es la variable confusora que se debe incluir en el modelo, en un primer momento se estimará un modelo de regresión logística multivariante en el cual se incluyan todas las variables confusoras (es lo que se conoce como modelo máximo). Mediante el método de selección de variables *backward*, se irán eliminando del modelo aquellas variables confusoras que no presenten significación estadística, considerando un nivel de significación estadística de $\alpha = 0.05$.

Después de haber estimado los diferentes modelos de regresión multivariable, se ha concluido que la variable confusora con menor p-valor asociado es el sexo.

Por lo tanto, el modelo de regresión logística final es el siguiente:

logistic Delirium NMB Sex

Logistic regression	Number of obs	=	538
	LR chi2 (2)	=	16.39
	Prob > chi2	=	0.0003
Log likelihood = -95.942025	Pseudo R2	=	0.0787

Delirium	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
NMB	3.181909	1.311208	2.81	0.005	1.418806	7.135964
Sex	4.055679	1.94386	2.92	0.003	1.585216	10.37621
_cons	.0033418	.0030338	-6.28	0.000	.0005639	.0198032

La estimación del OR obtenida mediante este modelo es:

$$OR_{multiv} = 3.18, \text{ con } IC_{95}(OR_{multiv}) = [1.42, 7.14]$$

Este resultado indica que, según este método de estimación del efecto, es 3.18 veces más probable que un paciente que ha recibido bloqueantes neuromusculares presente delirio frente a un paciente que no ha recibido el tratamiento.

4.4. Conclusiones sobre el escenario clínico I

Una vez que se ha estimado el efecto mediante los distintos métodos descritos en la teoría, se concluyen los siguientes aspectos:

Desde el punto de vista metodológico, se ha observado que los diferentes métodos del índice de propensión proporcionan estimaciones diferentes entre ellos, así como diferencias con el método clásico de regresión logística. Los resultados principales obtenidos se pueden observar en la *Tabla 6*:

Aplicaciones del Propensity Score								Sin Propensity Score			
Emparejamiento		IPTW		Estratificación MH		Covariable		RL Univariable		RL Multivariable	
OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)
1.73	(0.55, 5.97)	2.08	(0.83, 5.24)	1.82	(0.74, 4.42)	2.07	(0.87, 5.11)	2.79	(1.26, 6.17)	3.18	(1.42, 7.13)

Tabla 6 Resumen de los distintos métodos utilizado para el escenario clínico I

Cuando el número de eventos es reducido, así como el número de exposiciones, la estimación del efecto difiere entre métodos. dentro de los cuatro métodos del IP se han obtenido estimaciones similares entre el IPTW y el ajuste por la covariable, mientras que los OR obtenidos mediante el emparejamiento y de estratificación son más parecidos entre ellos.

Por lo que respecta a la estimación clásica, ésta obtiene valores más altos del OR, tanto de modo univariable como multivariable.

Simulación de modelos de emparejamiento

De modo orientativo, se ha realizado una simulación de 100 modelos para estudiar si influye el orden de los individuos cuando hay pacientes que presentan el mismo valor del IP estimado a la hora de realizar el emparejamiento.

Se ha observado que, cuando el número de variables que se incluyen en el IP es elevado, el orden de los sujetos no varía. Esto indica que no hay diferencia entre los sujetos y las parejas que se generan son siempre las mismas.

Por otro lado, también se ha observado que cuando se estima el IP sólo con las variables estadísticamente significativas, sí hay cambios en las simulaciones, pero cuando se incluyen más variables (no sólo las estadísticamente significativas si no también las clínicamente relevantes) entonces no hay cambios en las simulaciones.

En caso de que hubiera pacientes con valores del índice de propensión idénticos, el emparejamiento dependerá del número de variables que se incluyen en el modelo y de la significación estadística de las mismas.

5. Aplicación sobre un escenario clínico II

5.1. Búsqueda de variables confusoras

En esta segunda parte práctica, se quiere estudiar el efecto que provoca la interrupción de la sedación sobre la mortalidad de los pacientes ingresados en Unidades de Cuidados Intensivos.

La variable que se utilizará para medir la mortalidad es *Status_day28*, es una variable dicotómica que toma el valor 0 si la persona está viva al salir de la UCI y el valor 1 si la persona ha fallecido. Por lo tanto, el evento que se quiere estudiar es *Status_day28* = 1.

Por otro lado, la exposición será la interrupción de la sedación. La variable es *Daily_Awakening* que toma el valor 1 si al paciente le han interrumpido la sedación y el valor 0 si no le han interrumpido la sedación.

Si se quisiera estudiar la asociación que presenta la exposición con el evento, se obtendría el siguiente contraste de asociación Chi-cuadrado:

```
tab Daily_Awakening Status_day28, chi
```

see notes	Status_day28		Total
	Alive	Dead	
No	942	385	1,327
Yes	542	143	685
Total	1,484	528	2,012

```
Pearson chi2(1) = 15.4539 Pr = 0.000
```

Se observa que el p-valor asociado (0.000) indica una asociación estadísticamente significativa entre la interrupción de la sedación y la posterior mortalidad del paciente.

Considerando unas posibles variables confusoras, el gráfico causal del modelo es el que aparece en la *Figura 8*:

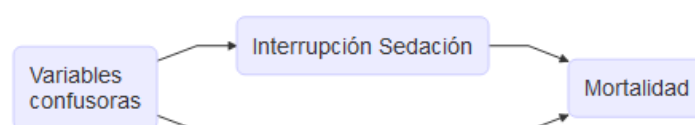


Figura 8 Relación causal entre la interrupción de la sedación, mortalidad y variables confusoras

Donde las variables confusoras presentan asociación tanto con la interrupción de la sedación como con la mortalidad.

Como ocurría en el escenario clínico I, para resolver los problemas de causalidad, el primer paso es buscar las posibles variables confusoras del modelo.

Mediante las *Tablas 7 y 8* se estudia la asociación de las variables basales con la exposición y con el evento:

- **Tabla de estudio de variables confusoras para la mortalidad:**

	Status_day28 Vivo (n=1677)	Status_day28 Fallecido (n=658)	P-valor
Edad	60,33 ± 16,30	66,187 ± 15,09	0,000
Sexo (Masculino)	850 (50,69)	363 (55,17)	0,051
IMC	27,33 ± 5,80	27,02 ± 5,56	0,082
Traqueotomía previa (Si)	41 (2,44)	13 (1,98)	0,497
SAPS_II	43,79 ± 17,05	57,16 ± 18,64	0,000
First Ventilatory Support (Non-invasive)	91 (5,43)	56 (8,51)	0,006
COPD (Si)	69 (4,14)	18 (2,74)	0,109
Asthma (Si)	12 (0,72)	2 (0,30)	0,243
CPD_NonCOPD (Si)	19 (1,14)	10 (1,52)	0,456
ARDS (Si)	47 (2,82)	23 (3,50)	0,388
Postoperative (Si)	355 (21,31)	72 (10,96)	0,000
CHF (Si)	111 (6,66)	45 (6,85)	0,871
Aspiration (Si)	39 (2,34)	11 (1,67)	0,319
Pneumonia (Si)	131 (7,86)	56 (8,52)	0,598
Community_pneumonia (Si)	98 (5,88)	35 (5,33)	0,604
Hospital_pneumonia (Si)	33 (1,98)	21 (3,20)	0,080
Sepsis (Si)	143 (8,83)	81 (12,60)	0,007
Trauma (Si)	74 (4,44)	11 (1,67)	0,001
Cardiac_arrest (Si)	76 (4,69)	86 (13,37)	0,000
Neurologic_disease (Si)	488 (29,29)	205 (31,20)	0,365
Coma_metabolic (Si)	96 (6,10)	24 (3,99)	0,055
ICU_pneumonia (Si)	72 (4,66)	27 (4,49)	0,864
sepsis_vm (Si)	254 (16,56)	121 (20,97)	0,018
cardiovascular_failure (Si)	648 (38,64)	314 (47,72)	0,000
renal_failure (Si)	354 (21,11)	195 (29,64)	0,000
hepatic_failure (Si)	414 (24,69)	145 (22,04)	0,177
hematologic_failure (Si)	183 (10,91)	99 (15,05)	0,006
barotrauma (Si)	23 (1,37)	13 (1,98)	0,286

	Status_day28 Vivo (n=1677)	Status_day28 Fallecido (n=658)	P-valor
RR_total_day1	18,07 ± 4,20	18,91 ± 5,00	0,002
respiratoryRateVent1	17,87 ± 4,12	18,06 ± 4,47	0,592
Tidal_Volume_day1	509,02 ± 85,30	499,99 ± 86,67	0,002
Tidal_mlgkg_day1	6,83 ± 1,59	6,86 ± 1,53	0,488
Tidal_mlgkgPBW_day1	8,31 ± 1,60	8,26 ± 1,66	0,350
peakPressure_day1	24,99 ± 7,37	26,24 ± 7,10	0,000
plateauPressure_day1	18,25 ± 5,67	19,72 ± 6,05	0,000
appliedPEEP_day1	6,31 ± 2,55	6,54 ± 2,60	0,069
driving_pressure_day1	12,06 ± 5,08	13,13 ± 5,48	0,003

Tabla 7 Estudio de las variables confusoras para el evento en el escenario clínico II

- **Tabla de estudio de variables confusoras para la exposición:**

	No interrupción Sedación (n=1.327)	Interrupción Sedación (n=685)	P-valor
Edad	61,44 ± 16,33	61,26 ± 15,65	0,56
Sexo (Masculino)	701 (52,83)	337 (49,20)	0,12
IMC	27,26 ± 5,63	27,28 ± 6,11	0,76
Traqueotomía previa (Si)	24 (1,81)	12 (1,75)	0,93
SAPS_II	47,45 ± 18,42	46,13 ± 17,42	0,20
First Ventilatory Support (Non-invasive)	87 (6,56)	39 (5,69)	0,45
COPD (Si)	57 (4,30)	21 (3,07)	0,18
Asthma (Si)	6 (0,45)	7 (1,02)	0,13
CPD_NonCOPD (Si)	16 (1,21)	12 (1,75)	0,32
ARDS (Si)	42 (3,17)	26(3,80)	0,46
Postoperative (Si)	230 (17,35)	140 (20,44)	0,09
CHF (Si)	99 (7,47)	38 (5,55)	0,11
Aspiration (Si)	35 (2,64)	15 (2,19)	0,54
Pneumonia (Si)	125 (9,43)	51 (7,45)	0,14
Community_pneumonia (Si)	92 (6,94)	35 (5,11)	0,11
Hospital_pneumonia (Si)	33 (2,49)	16 (2,34)	0,83
Sepsis (Si)	145 (11,21)	67 (10,09)	0,45
Trauma (Si)	54 (4,07)	27 (3,94)	0,89
Cardiac_arrest (Si)	76 (5,88)	49 (7,38)	0,20
Neurologic_disease (Si)	354 (26,70)	201 (29,34)	0,21
Coma_metabolic (Si)	59(4,76)	43 (6,66)	0,08
ICU_pneumonia (Si)	52 (4,33)	44 (6,94)	0,02
sepsis_vm (Si)	224 (18,95)	136 (22,01)	0,12

	No interrupción Sedación (n=1.327)	Interrupción Sedación (n=685)	P-valor
cardiovascular_failure (Si)	579 (43,63)	326 (47,59)	0,09
renal_failure (Si)	334 (25,17)	173 (25,26)	0,97
hepatic_failure (Si)	316 (23,81)	210 (30,66)	0,00
hematologic_failure (Si)	159 (11,98)	103 (15,04)	0,05
barotrauma (Si)	25 (1,88)	8 (1,17)	0,23
RR_total_day1	18,27 ± 4,46	18,49 ± 4,19	0,16
respiratoryRateVent1	18,17 ± 4,43	17,97 ± 4,15	0,42
Tidal_Volume_day1	511 ± 83,07	502,59 ± 86,88	0,04
Tidal_mlgkg_day1	6,87 ± 1,53	6,79 ± 1,66	0,19
Tidal_mlgkgPBW_day1	8,35 ± 1,59	8,21 ± 1,65	0,06
peakPressure_day1	25,77 ± 7,43	25,05 ± 7,13	0,07
plateauPressure_day1	18,61 ± 6,20	19,00 ± 51,14	0,11
appliedPEEP_day1	6,48 ± 2,72	6,31 ± 2,45	0,57
driving_pressure_day1	12,24 ± 5,49	12,71 ± 4,79	0,17

Tabla 8 Estudio de las variables confusoras para el tratamiento en el escenario clínico II.

Las variables que presentan relación con el evento son: edad, sexo, Saps_II, First_ventilatory_support, Post-operative, Sepsis, Trauma, Cardiac_arrest, Cardiovascular_failure, renal_failure, hematologic_failure, RR_total_day, Tidal_Volumeday1, peak_pressureday1, plateauPressure_day1.

Por lo tanto, están serán las variables confusoras que se utilizarán para la estimación del índice de Propensión.

5.2. Estimación y métodos del Índice de Propensión (PS)

5.2.1. Emparejamiento

Como en el escenario anterior, lo primero que se debe hacer es estimar el índice de propensión con las variables confusoras que se han obtenido en el apartado 5.1.

Mediante el comando **psmatch2** de Stata se realizan los siguientes cálculos:

```
psmatch2 Daily_Awakening age Sex SAPS_II First_ventilatory_support Postoperative
Sepsis Trauma Cardiac_arrest cardiovascular_failure renal_failure
hematologic_failure RR_total_day1 Tidal_Volume_day1 peakPressure_day1,
outcome(Status_day28) caliper (0.016) neighbor(1) noreplacement logit
```



```

Logistic regression
Number of obs      =      1,709
LR chi2(14)        =      33.61
Prob > chi2        =      0.0023
Pseudo R2         =      0.0151
Log likelihood = -1092.6527

```

Daily_Awakening	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
age	.0028521	.003385	0.84	0.399	-.0037823	.0094866
Sex	.1559267	.1033186	1.51	0.131	-.046574	.3584274
SAPS_II	-.0051356	.0032915	-1.56	0.119	-.0115868	.0013157
First_ventilatory_support	-.2768774	.3418172	-0.81	0.418	-.9468268	.3930721
Postoperative	.246323	.1375624	1.79	0.073	-.0232944	.5159403
Sepsis	-.0575988	.180437	-0.32	0.750	-.4112487	.2960512
Trauma	.0381411	.2590416	0.15	0.883	-.4695712	.5458534
Cardiac_arrest	.5847191	.2169783	2.69	0.007	.1594494	1.009989
cardiovascular_failure	.1347309	.1107868	1.22	0.224	-.0824072	.3518691
renal_failure	-.1937562	.1432419	-1.35	0.176	-.4745051	.0869927
hematologic_failure	.3921784	.170133	2.31	0.021	.0587239	.7256329
RR_total_day1	.0172522	.0127963	1.35	0.178	-.007828	.0423325
Tidal_Volume_day1	-.0013563	.0006612	-2.05	0.040	-.0026522	-.0000605
peakPressure_day1	-.0148121	.0073856	-2.01	0.045	-.0292876	-.0003366
_cons	-.1669588	.5504549	-0.30	0.762	-1.245831	.9119129

Se estima el índice de propensión mediante un modelo de regresión logística donde la probabilidad que se modeliza es la de que, a un paciente, en función de sus condiciones basales, le hayan interrumpido o no la sedación.

Una vez que se ha estimado el IP, el *psmatch2* realiza el emparejamiento de forma instantánea:

Variable	Sample	Treated	Controls	Difference	S.E.	T-stat
Status_day28	Unmatched	.205638474	.28119349	-.075555016	.021987683	-3.44
	ATT	.206666667	.29	-.083333333	.024848642	-3.35

Note: S.E. does not take into account that the propensity score is estimated.

Se observa que el efecto que produce la interrupción de la sedación es muy similar antes y después del emparejamiento. Además, los resultados que se obtiene entre tratados y no tratados son muy similares entre ellos en los dos momentos.

Hay un efecto mayor de la mortalidad en aquellos pacientes que no han interrumpido la sedación. Se podría considerar por lo tanto que, pacientes que interrumpen la sedación presentan un riesgo menor de fallecer.

psmatch2:	psmatch2: Common		
Treatment	support		
assignment	Off suppo	On suppor	Total
-----+-----+-----			
Untreated	0	1,106	1,106
Treated	3	600	603
-----+-----+-----			
Total	3	1,706	1,709

Considerando la zona de soporte común se observa que en este escenario clínico casi todos los pacientes forman parte del emparejamiento. De los 1709 pacientes que se estudian, sólo se quedan fuera del emparejamiento 3 de ellos.

Esto significa que la zona de soporte común abarca prácticamente la totalidad de los pacientes, tanto tratados y no tratados. Mediante la *Figura 9* donde se representan los gráficos de cajas del IP para tratados y no tratados, se observa que la distribución en ambos subgrupos es muy similar:

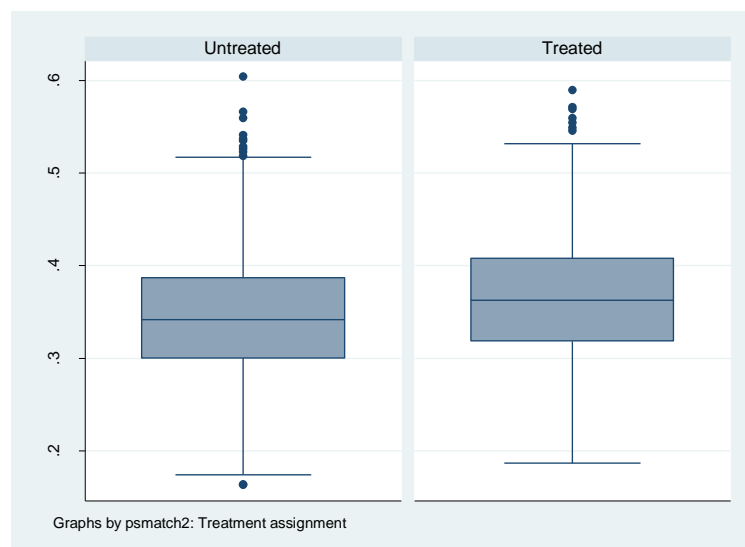


Figura 9 Zona de soporte común en el Escenario Clínico II

El efecto que produce la interrupción de la sedación sobre la mortalidad en pacientes ingresados en UCI es de:

$$OR_{match} = 0.6377, IC_{95}(OR_{match}) = [0.48, 0.84]$$

La interrupción de la sedación reduce en un 37% la probabilidad de fallecer en pacientes ingresados en UCI.

El OR obtenido está asociado con un p-valor de 0.0008, lo que indica que este resultado es estadísticamente significativo. Además, el Intervalo de confianza al 95% para el OR no contiene el 1, lo que en la práctica habitual indica que es un valor significativo.

Para obtener este valor, se ha utilizado el siguiente comando de Stata:

```
cc Status_day28 Daily_Awakening if _weight ==1
```

	Exposed	Unexposed	Total	Proportion Exposed
Cases	124	174	298	0.4161
Controls	476	426	902	0.5277
Total	600	600	1200	0.5000
	Point estimate		[95% Conf. Interval]	
Odds ratio	.6377861		.4847508	.8381739 (exact)
Prev. frac. ex.	.3622139		.1618261	.5152492 (exact)
Prev. frac. pop	.1911461			
chi2(1) = 11.16 Pr>chi2 = 0.0008				

Como parte del método del emparejamiento, se debe comprobar que la diferencia en entre tratados y no tratados antes y después del emparejamiento sea mínima. Como se ha explicado en el apartado 2.2.1. se deben calcular las diferencias estandarizadas de las variables confusoras del modelo.

En la *Tabla 9*, se muestran los resultados obtenidos para cada variable basal en cada categoría del tratamiento, antes y después del emparejamiento. La diferencia obtenida debe ser menor que un 10% para que efectivamente ambos grupos no presenten diferencias. En caso contrario, significará que el índice de propensión no está bien indicado.

Variables basales	Antes del emparejamiento			Después del emparejamiento		
	Daily_Awakening = Si (n=685)	Daily_Awakening = No (n=1327)	Diferencias estandarizadas (%)	Daily_Awakening = Si (n=602)	Daily_Awakening = No (n=602)	Diferencias estandarizadas (%)
Edad (años)	61,26 ± 15,66	61,44 ± 16,34	1,09	61,36 ± 15,78	60,98 ± 16,26	2,38
SAPS_II	46,13 ± 17,42	47,45 ± 18,42	7,37	46,39 ± 17,70	46,38 ± 19,02	0,054
RR_total_day1	18,49 ± 4,19	18,27 ± 4,46	4,98	18,44 ± 4,11	18,54 ± 4,57	2,33
Tidal_Volume_day1	502,59 ± 86,88	511,88 ± 83,07	10,93	501,01 ± 84,07	501,62 ± 74,10	0,77
peakPressure_day1	25,05 ± 7,13	25,77 ± 7,43	9,96	25,02 ± 7,12	25,17 ± 7,12	2,17
Sexo (V)	49,2	52,83	7,26	49,34	47,34	4,00
First_ventilatory_support (Si)	94,31	93,44	1,31	97,84	98,17	2,36
Postoperative (Si)	20,44	17,35	5,25	21,2	20,43	1,90
Sepsis (Si)	10,09	11,21	1,76	89,87	89,87	0,00
Trauma (Si)	3,94	4,07	0,19	4,15	4,15	0,00
Cardiac_arrest (Si)	7,38	5,88	2,27	7,97	7,81	0,59
cardiovascular_failure (Si)	47,59	43,63	7,89	48,17	48,17	0,00
renal_failure (Si)	25,26	25,17	0,16	25,42	24,75	1,55
hematologic_failure (Si)	15,04	11,98	4,95	14,95	15,28	0,92

Tabla 9. Comparación de las variables basales entre tratados y no tratados antes y después del emparejamiento en el escenario clínico II

Al estudiar las diferencias estandarizadas se concluye que las distribuciones de los pacientes tratados y no tratados después del emparejamiento son muy similares ya que ninguna variable basal presenta una diferencia mayor que el 10%.

Este suceso indica dos cosas:

- Por una parte, que el modelo de regresión logística que estima el índice de propensión está especificado correctamente puesto que las diferencias estandarizadas no superan el límite establecido.
- Además, al no haber diferencias antes del emparejamiento, indica que, desde el principio, ambos grupos de pacientes no presentan diferencias en cuanto a las características basales entre ellos.

5.2.2. IPTW

A continuación, se estima el efecto de la interrupción de la sedación ponderando por la inversa de los pesos del índice de propensión.

Como se ha explicado en el apartado 4.2.2., para obtener la estimación del efecto con este método IPTW, lo primero que se debe hacer es calcular los pesos de los pacientes que luego servirán para la estimación.

```
gen iptw = 1 / _pscore if _treated==1
replace iptw= 1 / (1- _pscore) if _treated == 0
```

Mediante el comando **logistic** se estima el modelo:

```
logistic Status_day28 Daily_Awakening [pweight=iptw]
```

```
Logistic regression               Number of obs   =       1,709
                                Wald chi2(1)       =       10.74
                                Prob > chi2        =       0.0011
Log pseudolikelihood = -1887.0777 Pseudo R2       =       0.0067
```

		Robust				
Status_day28	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Daily_Awakening	.6684829	.0821673	-3.28	0.001	.5253684	.8505829
_cons	.3912072	.0263518	-13.93	0.000	.3428229	.4464202

Se observa que, con la interrupción de la sedación, se reduce un 33,15% el riesgo de fallecer frente a los que no interrumpen la sedación (OR= 0.67).

Este resultado es estadísticamente significativo puesto que el p-valor asociado al contraste es de 0.001 y el intervalo de confianza al 95% para el OR estimado no contiene el valor 1.

$$OR_{iptw} = 0.67 \text{ con } IC_{95}(OR_{iptw}) = [0.52, 0.85]$$

5.2.3. Estratificación

En este método lo primero que se debe hacer es calcular los valores de los quintiles del índice de propensión.

Los valores de los quintiles del PS son:

```
centile _pscore, centile(20 40 60 80)
```

Variable	Obs	Percentile	Centile	-- Binom. Interp. -- [95% Conf. Interval]	
_pscore	1,709	20	.296915	.2931068	.3008408
		40	.3335286	.3293693	.3372574
		60	.3643792	.3607886	.3693849
		80	.4063733	.4018841	.4113439

Una vez que se han calculado los quintiles, se genera la variable que conformará los estratos:

```
recode _pscore (min/.296915=1 "Estrato 1") (.296915/.3335286=2 "Estrato 2") (.3335286/.3643792 =3 "Estrato 3") (.3643792/.4063733 = 4 "Estrato 4") (.4063733 /max = 5 "Estrato 5") ,gen(ps_estr)
```

A partir de la nueva variable **ps_estr**, se calcula el efecto estimado de la exposición:

```
cc Daily_Awakening Status_day28, by(ps_estr )
```

Los resultados que se obtienen son los siguientes:

RECODE of _pscor	OR	[95% Conf. Interval]		M-H Weight
Estrato 1	.6300366	.3386891	1.137924	16.01173 (exact)
Estrato 2	.4986197	.2583458	.9256274	16.94737 (exact)
Estrato 3	.6319444	.3596098	1.091545	18.52632 (exact)
Estrato 4	1.180347	.6731378	2.053976	12.78717 (exact)
Estrato 5	.5806452	.3363831	.991852	20 (exact)
Crude	.6617484	.5175376	.8434841	(exact)
M-H combined	.6758078	.5319384	.8585884	
Test of homogeneity (M-H) chi2(4) = 5.84 Pr>chi2 = 0.2116				
Test that combined OR = 1:				
		Mantel-Haenszel	chi2(1) =	10.36
			Pr>chi2 =	0.0013

Se observa que los valores de los OR obtenidos concuerdan con los que se obtenían en los métodos anteriores.

$$OR_{crudo} = 0.66, IC_{95}(OR_{crudo}) = [0.52, 0.84]$$

$$OR_{mantel} = 0.68, IC_{95}(OR_{mantel}) = [0.53, 0.86]$$

Ambos OR presentan un intervalo de confianza estrecho. Éstos, al no contener el valor 1, están indicando que los valores que se han estimado de los OR son estadísticamente significativos.

Se puede concluir por lo tanto que la interrupción de la sedación reduce, en un 32 % la probabilidad de fallecer, en pacientes ingresados en UCIs.

5.2.4. Incluyendo el PS como covariable del modelo

En este último método de aplicación del IP se ha utilizado el índice de propensión como si fuera una covariable más del modelo para estimar el efecto que produce la interrupción de la sedación en la mortalidad.

Como se ha realizado en el apartado 4.2.4. se ha incluido el índice de propensión como covariable del modelo de regresión logística:

```
logistic Status_day28 Daily_Awakening _pscore
```

```
Logistic regression               Number of obs   =      1,709
                                LR chi2(2)         =      12.34
                                Prob > chi2         =      0.0021
Log likelihood = -963.28265       Pseudo R2      =      0.0064
```

Status_day28	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Daily_Awakening	.6679276	.0815005	-3.31	0.001	.5258547	.8483851
_pscore	.6176656	.5215946	-0.57	0.568	.1180176	3.23266
_cons	.4620517	.1381359	-2.58	0.010	.257166	.8301708

El valor del OR estimado es:

$$OR_{covariable} = 0.67, IC_{95}(OR_{covariable}) = [0.52, 0.85]$$

5.3. Estimación del efecto sin el índice de propensión

Si se quisiera ver cuál sería el efecto de la interrupción de la sedación sin el índice de propensión, se estimaría un modelo de regresión logística:

```
logistic Status_day28 Daily_Awakening
```

```

Logistic regression                                Number of obs   =      2,012
                                                    LR chi2(1)      =      15.85
                                                    Prob > chi2     =      0.0001
Log likelihood = -1150.1384                      Pseudo R2       =      0.0068

```

Status_day28	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Daily_Awakening	.6455456	.0721655	-3.92	0.000	.5185267	.8036792
_cons	.4087049	.0247223	-14.79	0.000	.3630121	.4601491

El OR que se obtiene al estimar mediante regresión logística (sin índice de propensión) es muy similar al que se obtenía cuando se consideraba el propensity score. La interrupción de la sedación reduce un 35,44% la probabilidad de fallecer.

Si se incluyeran en el modelo todas las variables que se han utilizado para el índice de propensión:

```

Logistic regression                                Number of obs   =      1,709
                                                    LR chi2(15)     =     237.07
                                                    Prob > chi2     =      0.0000
Log likelihood = -850.91668                      Pseudo R2       =      0.1223

```

Status_day28	Odds Ratio	Std. Err.	z	P> z	[95% Conf. Interval]	
Daily_Awakening	.6297881	.0826875	-3.52	0.000	.4868964	.8146148
age	1.019964	.0043406	4.65	0.000	1.011492	1.028507
Sex	.7960444	.0971461	-1.87	0.062	.6267004	1.011148
SAPS_II	1.030568	.0038406	8.08	0.000	1.023068	1.038123
First_vent_supp	1.707535	.5679818	1.61	0.108	.8896765	3.27723
Postoperative	.5617854	.1041104	-3.11	0.002	.3906837	.8078219
Sepsis	1.007218	.1925952	0.04	0.970	.6924058	1.465163
Trauma	.6542086	.2405312	-1.15	0.248	.3182418	1.344854
Cardiac_arrest	1.973929	.4597051	2.92	0.004	1.250537	3.115778
cardiova_failure	1.347283	.1753701	2.29	0.022	1.043907	1.738826
renal_failure	.9564342	.1485974	-0.29	0.774	.7053542	1.296889
hemat_failure	1.354951	.2544516	1.62	0.106	.9377197	1.957826
RR_total_day1	1.018951	.014714	1.30	0.194	.9905163	1.048202
Tidal_Vol_day1	.9995693	.0007586	-0.57	0.570	.9980836	1.001057
peakPres_day1	1.009385	.0086647	1.09	0.277	.9925445	1.026511
_cons	.0214448	.0142302	-5.79	0.000	.005841	.078733

Los resultados que se obtienen son similares, aunque en este caso la probabilidad de riesgo toma un valor más alto que en el modelo anterior ($1 - 0.6297 = 0.3703$).

5.4. Conclusiones sobre el escenario clínico II

Considerando la *Tabla 10*, las conclusiones que se obtiene acerca del escenario clínico II son las siguientes:

- La interrupción de la sedación reduce la probabilidad de fallecer en pacientes ingresados en la UCI (OR=0.64, IC 95% 0.49-0.83).
- Al utilizar los distintos métodos de aplicación que presenta el índice de propensión las estimaciones obtenidas no difieren entre ellas. Los valores de los OR y sus IC al 95 % son muy similares.
- Además, al comparar los métodos del IP con la regresión logística clásica se obtienen estimaciones del efecto muy similar.
- Observando la *Figura 9*, se concluye que el comportamiento que presentan tanto pacientes tratados y no tratados es muy parecido.
- Por eso mismo, se podría concluir que la interrupción de la sedación no se aplica a un conjunto específico de pacientes.

Aplicaciones del Propensity Score								Sin Propensity Score			
Emparejamiento		IPTW		Estratificación MH		Covariable		RL Univariable		RL Multivariable	
OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)	OR	IC (95%)
0,64	(0,49; 0,83)	0,67	(0,52; 0,85)	0,68	(0,53; 0,86)	0,67	(0,52; 0,85)	0,64	(0,52; 0,80)	0,63	(0,49; 0,81)

Tabla 10 Resumen de los distintos métodos utilizados en el escenario clínico II

6. Conclusiones

Después de haber estudiado distintos escenarios clínicos en los cuales la prevalencia de la exposición y de los eventos es distinta, se pueden concluir los siguientes aspectos:

- En el escenario I, donde el número de eventos y de exposiciones era reducida, los diferentes métodos de estimación de la causalidad producían resultados, en términos de OR y de Intervalos de confianza, diferentes.
- Si los pacientes expuestos y no expuestos difieren, hay diferencias entre las estimaciones que se obtienen mediante el Índice de Propensión y la regresión logística clásica.
- En el escenario II, donde pacientes tratados y no tratados no presentan diferencias estadísticamente significativas en cuanto a sus características basales, la estimación de la causalidad obtenida mediante el IP y la regresión logística es muy similar.
- Dependiendo del número de eventos y de exposiciones que haya, el índice de propensión y la regresión logística podrían considerarse equivalentes, obteniendo valores en las estimaciones similares cuando no se observan diferencias entre pacientes tratados y no tratados y el número de eventos es elevado.
- Mediante el método del emparejamiento del índice de propensión, se obtiene una estructura semejante a la de un ensayo clínico, con la desventaja de que aquellos pacientes que no encuentren pareja serán descartados del estudio.
- Cuando el número de variables confusoras es elevado y el número de eventos es reducido, el índice de propensión permite ajustar por más variables que el modelo clásico de regresión logística.
- Las características del escenario teórico subyacente deben ser tenidas en cuenta para seleccionar el método más adecuado de ajuste en los análisis de causalidad. El índice de propensión permite seleccionar adecuadamente variables por las que corregir.

- Si existen diferencias en los perfiles de exposición, la adecuación de los distintos métodos depende de la prevalencia de las exposiciones y de la incidencia de los desenlaces.

7. Bibliografía

1. Laporte JR. Principios básicos de investigación clínica. Barcelona. Astra Zeneca. 2001.
2. Austin, Peter C. and Laupacis, Andreas (2011) A Tutorial on Methods to Estimating Clinically and Policy-Meaningful Measures of Treatment Effects in Prospective Observational Studies: A Review. The International Journal of Biostatistics: Vol. 7 : Iss. 1, Article 6.
3. Rosenbaum PR, Rubin D. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 1983;70(1):41-55.
4. Kahlert J, Gribsholt SB, Gammelager H, Dekkers OM, Luta G. Control of confounding in the analysis phase- an overview for clinicians. *Clin Epidemiol* 2017; (9): 195–204
5. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011 May;46(3):399-424.
6. Austin PC. An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies. *Multivariate Behav Res* 2011 May;46(3):399-424.
7. Guo S, Fraser MW. Propensity Score Analysis Statistical Methods and Applications 1. SAGE. 2009
8. Austin PC. A Tutorial and Case Study in Propensity Score Analysis: An Application to Estimating the Effect of In-Hospital Smoking Cessation Counseling on Mortality *Multivariate Behav Res* 2011; 46(1):119-151.
9. Patrick AR, Schneeweiss S, Brookhart MA, Glynn RJ, Rothman KJ, Avorn J, et al. The implications of propensity score variable selection strategies in pharmacoepidemiology: an empirical illustration. *Pharmacoepidemiol Drug Saf* 2011 Jun;20(6):551-9
10. Brookhart MA, Schneeweiss S, Rothman KJ, Glynn RJ, Avorn J, Sturmer T. Variable selection for propensity score models. *Am J Epidemiol* 2006 Jun 15;163(12):1149-56

11. Deb S., Austin PC. A Review of Propensity-Score Methods and Their Use in Cardiovascular Research. Canadian Journal of Cardiology (2016); Multivariate Behav Res 2011; 32 (1): 259-265.
12. Peduzzi P, Concato J, Kemper E, Holford TR, Feinstein AR. A simulation study of the number of events per variable in logistic regression analysis. J Clin Epidemiol 1996 Dec;49(12):1373-9.
13. Harrell FE, Jr., Lee KL, Mark DB. Multivariable prognostic models: issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Stat Med 1996 Feb 28;15(4):361-87
14. Cepeda MS, Boston R, Farrar JT, Strom BL. Comparison of Logistic Regression versus Propensity Score When the Number of Events Is Low and There Are Multiple Confounders. Am J Epidemiol 2003 Aug 1;158(3):280.